

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICA**

Vinícius Sousa Fazio

OTIMIZAÇÃO MULTIOBJETIVO

Florianópolis

2018

Vinícius Sousa Fazio

OTIMIZAÇÃO MULTIOBJETIVO

Trabalho de Conclusão de Curso submetido ao Departamento de Matemática do Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina para a obtenção do grau de Bacharel em Matemática.

Orientador:
Prof. Dr. Douglas Soares Gonçalves

Florianópolis

2018

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Fazio, Vinícius Sousa
Otimização Multiobjetivo / Vinícius Sousa Fazio ;
orientador, Douglas Soares Gonçalves, 2018.
138 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro de
Ciências Físicas e Matemáticas, Graduação em
Matemática, Florianópolis, 2018.

Inclui referências.

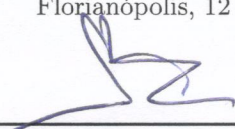
1. Matemática. 2. Otimização Multiobjetivo. 3.
Método de Descida. 4. Método de Cauchy
Multiobjetivo. I. Gonçalves, Douglas Soares. II.
Universidade Federal de Santa Catarina. Graduação em
Matemática. III. Título.

Vinícius Sousa Fazio

OTIMIZAÇÃO MULTIOBJETIVO


Este Trabalho de Conclusão de Curso foi julgado aprovado para a obtenção do título de “Bacharel em Matemática”, e aprovado em sua forma final pelo Departamento de Matemática do Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina.

Florianópolis, 12 de novembro 2018.



Prof.^a Dr.^a Sonia Elena Palomino Castro
Coordenadora do Curso


Banca Examinadora:



Prof. Dr. Douglas Soares Gonçalves
Orientador



Prof. Dr. Maicon Marques Alves



Prof. Dr. Juliano de Bem Francisco

RESUMO

A otimização multiobjetivo consiste em minimizar simultaneamente diversas funções escalares sujeitas a restrições. Para isso, é definida uma relação de ordem parcial e usado o conceito de *ótimo de Pareto* em vez de minimizador. Existem diversos métodos para encontrar ótimos de Pareto, como os *métodos de descida*, que são métodos iterativos. São apresentados quatro métodos de descida, entre os quais dois são contribuição deste trabalho. É proposta uma generalização das *sequências gradient-related*, que é um conceito usado para demonstrar que, sob certas condições, as sequências geradas pelos métodos de descida convergem para pontos críticos, no contexto de minimização escalar. A generalização proposta visa garantir a convergência a pontos críticos, no contexto multiobjetivo. Os métodos apresentados também são comparados em alguns experimentos numéricos, como o problema de portfólio.

Palavras-chave: Otimização Multiobjetivo. Método de Descida. Método de Cauchy Multiobjetivo.

ABSTRACT

Multi-objective optimization consists of minimizing simultaneously several scalar functions subjected to constraints. For this purpose, a partial order relation is defined and the concept of *Pareto optimal* solution is used instead of minimizer. There are several methods to find a Pareto optimal solution, such as *descent methods*, which are iterative methods. Four iterative descent methods are presented, among which two of them are contributions of this work. A generalization of the *gradient-related sequences* is proposed, which is a concept used to demonstrate that, under certain conditions, the sequences generated by the descent methods converge to critical points in the context of scalar minimization. The proposed generalization aims to guarantee the convergence to critical points, in the multi-objective context. The presented methods are also compared in some numerical experiments, such as the portfolio problem.

Keywords: Multi-objective Optimization. Descent Method. Multi-objective Steepest Descent Method.

Sumário

	Introdução	7
1	OTIMIZAÇÃO MULTIOBJETIVO	15
1.1	Ótimo de Pareto e ótimo fraco de Pareto . .	15
1.2	Direção factível de descida e ponto crítico .	18
2	CONVERGÊNCIA DE MÉTODOS DE DESCIDA MULTIOBJETIVO	27
2.1	Método de descida	27
2.2	Método de descida com backtracking	30
2.3	Aplicação regular de direções	31
2.3.1	Convergência em funções continuamente diferenciáveis	32
2.3.2	Convergência em funções convexas diferenciáveis	39
2.3.3	Algoritmo do método de descida com aplicação regular de direções	50
3	DIREÇÃO DE CAUCHY MULTIOBJETIVO	53
3.1	Análise da direção de Cauchy multiobjetivo	54
3.2	Exemplos de uso	70
4	OUTRAS DIREÇÕES DE DESCIDA	75
4.1	Direção Central	75
4.2	Direção Linear	82
4.3	Direção de Pareto descendente	88
5	EXPERIMENTOS NUMÉRICOS	97

5.1	Coleção de problemas	98
5.1.1	Problemas de funções convexas	99
5.1.2	Problemas de funções não convexas	102
5.1.3	Problemas difíceis	107
5.2	Portifólio	108
	Conclusão	113
	REFERÊNCIAS	115
	APÊNDICE A – CONCEITOS BÁSICOS . .	119
	APÊNDICE B – FUNÇÕES CONVEXAS .	121
	APÊNDICE C – EXEMPLOS ADICIONAIS	125
	APÊNDICE D – FORMAS DIRETAS	131

Introdução

A *otimização escalar* é a área da matemática que desenvolve métodos para encontrar os mínimos de funções reais $f: \mathbb{R}^n \rightarrow \mathbb{R}$ restritas a um subconjunto do domínio $C \subseteq \mathbb{R}^n$. O conjunto C normalmente é chamado *conjunto viável* (NOCEDAL; WRIGHT, 1999).

A *otimização multiobjetivo* estende o conceito de otimização escalar, e consiste em minimizar simultaneamente as funções componentes $f_j: \mathbb{R}^n \rightarrow \mathbb{R}$, $\forall j \in \{1, \dots, m\}$, de uma função vetorial $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ restrita a um subconjunto $C \subseteq \mathbb{R}^n$ (GRAÑA DRUMMOND; SVAITER, 2015).

Porém, o conjunto \mathbb{R}^m não é totalmente ordenado para $m > 1$, e na maioria dos casos os objetivos são conflitantes. A seguinte ordem parcial é considerada:

$$u, v \in \mathbb{R}^m: \quad u \preceq v \iff u_i \leq v_i, \quad \forall i \in \{1, \dots, m\}.$$

Em um conjunto de ordem parcial, nem sempre um par de elementos é comparável. Por isso, no contexto de otimização multiobjetivo, em vez de *minimizador* utiliza-se o conceito de *ótimo de Pareto* (GRAÑA DRUMMOND; SVAITER, 2015), que podem ser vistos como minimizadores na ordem parcial do \mathbb{R}^m .

O objetivo da otimização multiobjetivo, portanto, é encontrar ótimos de Pareto.

Pergunta de pesquisa

Sob quais condições é possível encontrar ótimos de Pareto, e como encontrar tais pontos?

O escopo deste trabalho é analisar a otimização multiobjetivo em funções diferenciáveis e com conjunto viável convexo e fechado.

Justificativa

Métodos de otimização multiobjetivo têm diversas aplicações na indústria, pois diversos problemas podem ser interpretados como minimização de funções vetoriais. Exemplos:

- Projetar navios porta-contêineres ([PRIFTIS et al., 2018](#)). A otimização multiobjetivo é utilizada para minimizar simultaneamente o custo operacional, de aquisição, e provenientes de exigências legais, e ainda maximizar a capacidade.
- Usinas hidrotermoeletricas ([QIN et al., 2010](#)), onde se utiliza otimização multiobjetivo para reduzir o custo com combustível e ao mesmo tempo reduzir a emissão de gases poluentes.
- Separadores ciclônicos de gases, que são equipamentos utilizados na indústria para extrair partículas sólidas em suspensão num escoamento de gases([LUCIANO et al., 2018](#)). A otimização multiobjetivo é utilizada para maximizar a eficiência de coleta e ao mesmo tempo minimizar a queda de pressão, que são os parâmetros mais relevantes na avaliação de desempenho de separadores ciclônicos de gases.
- Pistão de motores a diesel ([DENG et al., 2018](#)). A otimização multiobjetivo junto com simulações são utilizadas para projetar o sistema de resfriamento de motores a diesel a fim de minimizar a temperatura máxima atingida pelo motor e o estresse térmico máximo ao qual o motor é submetido.

- Escalonamento de Processos (CHIANG; LIN, 2013). O problema de escalonamento de processos é definir a ordem em que as demandas serão atendidas, e a distribuição dos recursos disponíveis, a fim de minimizar simultaneamente o tempo médio de atendimento, o tempo máximo e o tempo total.

Métodos

Existem diversos métodos de otimização multiobjetivo, com diferentes condições sobre a função $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ e conjunto viável $C \subseteq \mathbb{R}^n$ que garantem que um ótimo de Pareto será encontrado. A seguir serão listados alguns métodos para uma visão geral, mas o foco deste trabalho são os *métodos de descida*.

Escalarização

Os métodos de *escalarização*, em geral, definem uma função $g: \mathbb{R}^m \rightarrow \mathbb{R}$, chamada *função escalarizadora*, que é aplicada sobre $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ (MIETTINEN; MÄKELÄ, 2002). Então o problema é reformulado para

$$\operatorname{argmin}_{x \in C \subseteq \mathbb{R}^n} g \circ f(x).$$

Este novo problema pode ser resolvido com métodos de otimização escalar. Diferentes métodos de escalarização usam funções g diferentes, e cada método possui suas condições para que a solução encontrada seja um ótimo de Pareto.

Um exemplo de função escalarizadora é o método da *soma ponderada* (MARLER; ARORA, 2004):

$$g \circ f(x) = \sum_{i=1}^m w_i f_i(x),$$

onde $w \in \mathbb{R}^m$ é um vetor de pesos não negativos. Neste caso, é necessário exigir certas condições para as funções componentes f_i , para os pesos w_i e para o conjunto $C \subseteq \mathbb{R}^n$ para que seja possível obter um ótimo de Pareto. Por exemplo:

$$\begin{aligned}(w_1, w_2) &= (1, 1) \\ (f_1(x), f_2(x)) &= (x^3, -x) \\ \min_{x \in \mathbb{R}} g \circ f(x) &= x^3 - x.\end{aligned}$$

Neste caso, $g \circ f(x)$ é ilimitada inferiormente, então não é possível minimizá-la. Porém, qualquer ponto $x \in \mathbb{R}$ é um ótimo de Pareto no problema original, pois f_1 é crescente e f_2 é decrescente.

Método Lexicográfico

O *método lexicográfico* consiste em minimizar uma função componente de cada vez, em ordem de prioridade (MARLER; ARORA, 2004). Supondo que as funções componentes f_i já estão em ordem de prioridade, o problema de otimização se transforma em m subproblemas de otimização escalar. Para cada $i \in \{1, \dots, m\}$,

$$\begin{aligned}x_i &:= \operatorname{argmin}_{x \in C \subseteq \mathbb{R}^n} f_i(x) \\ \text{sujeito a } f_j(x) &\leq f_j(x_{i-1}), \quad \forall j \in \{1, \dots, i-1\}.\end{aligned}$$

Neste caso, a solução do método lexicográfico é o x_m .

Normal Boundary Intersection (NBI)

No método Normal Boundary Intersection, primeiro encontra-se o mínimo globais de cada uma das m funções componentes (MARLER; ARORA, 2004; RIGONI; POLES, 2005). A partir

destes m pontos da imagem de f , define-se uma combinação convexa destes pontos. Em seguida, escolhe-se um vetor normal ao plano formado pelas combinações convexas. Enfim, o problema é reduzido a minimização da função f restrita ao vetor normal escolhido, que é um problema de uma otimização escalar.

Este método, sob certas hipóteses sobre a função f , possui a vantagem de facilmente determinar diversos ótimos de Pareto e espaçados uniformemente, o que é desejável para se obter uma boa amostra do conjunto de ótimos de Pareto. Porém, o conjunto $C \subseteq \mathbb{R}^n$ não pode ser qualquer, e a imagem de cada função componente no conjunto viável precisa ser limitada inferiormente.

Algoritmos genéticos

Nos algoritmos genéticos, geralmente se inicia com uma *população* (pontos do conjunto viável) (ZHOU et al., 2011). A cada *geração* (iteração) são feitos *cruzamentos* (uma combinação de pontos da população para gerar novos pontos) e *mutações* (ruídos adicionados em alguns pontos da população). Após várias gerações, espera-se obter uma população com diversos ótimos de Pareto. Estes métodos variam normalmente na forma de fazer o cruzamento, nas mutações e no critério de parada.

Métodos de descida

Os métodos de descida são o foco deste trabalho, e são analisados mais detalhadamente (GRAÑA DRUMMOND; SVAITER, 2015; FLIEGE; SVAITER, 2000).

Estes métodos se aplicam somente às funções diferenciáveis, pois o gradiente das funções componentes são utilizados.

Utiliza-se o conceito de *direção factível de descida* de um ponto $x \in \mathbb{R}^n$ (definido no Capítulo 1), que pode ser visto como uma direção $v \in \mathbb{R}^n$ que forma um ângulo de mais de 90° com cada gradiente de f em x . Um ponto sem direção factível de descida é um ponto crítico, e todo ponto ótimo de Pareto é um *ponto crítico*.

Um algoritmo para tentar encontrar um ponto crítico (definido no Capítulo 2) consiste em encontrar uma direção factível de descida, andar nesta direção, encontrar outra direção de descida, e repetir o processo até que um ponto crítico seja encontrado, o que gera uma sequência de pontos. Existem diferentes métodos para determinar direções factíveis de descida, e cada método tem suas próprias condições para garantir que um ótimo de Pareto será encontrado.

Uma das contribuições deste trabalho foi determinar condições suficientes sob as direções factíveis de descida para se garantir que todo ponto limite da sequência gerada pelo algoritmo seja um ponto crítico, inspirado nos trabalhos (FUKUDA; DRUMMOND, 2014) e (GRAÑA DRUMMOND; SVAITER, 2015). Isso é obtido com uma *aplicação regulares de direções* (definida na Seção 2.3), que pode ser visto como uma generalização das *sequências gradient-related* (BERTSEKAS, 1999), mas em vez de se restringir à otimização escalar, uma aplicação regular de direção é usada no contexto multiobjetivo.

Quatro métodos de direção factível de descida foram analisados (Capítulos 3 e 4).

A direção de Cauchy multiobjetivo (GRAÑA DRUMMOND; SVAITER, 2015) foi analisada, e uma variação foi proposta: a direção central. A direção de Cauchy multiobjetivo possui uma tendência de tomar direções de descida próximas ao gradiente

negativo de menor norma, enquanto a direção central tende a tomar direções de descida no centro do cone formado pelos gradientes negativos. Ambas direções são obtidas pela solução de um subproblema de programação quadrática.

A direção linear (FLIEGE; SVAITER, 2000) e uma variação proposta da direção de Pareto descendente de (HARADA et al., 2006) foram também analisadas, que são direções obtidas através da solução de um subproblema de programação linear.

Foi demonstrado que os quatro métodos analisados satisfazem as condições da aplicação regular de direções.

Uma condição um pouco mais forte, chamada de aplicação fortemente regular de direções, também foi proposta. Sob certas condições na função a ser minimizada, demonstrou-se que a sequência gerada pelo algoritmo utilizando uma aplicação fortemente regular de direções converge para um ponto crítico. Porém, apenas as direções de Cauchy multiobjetivo e central satisfazem as condições de aplicação fortemente regular de direções.

Todas direções analisadas são comparadas em alguns experimentos numéricos no Capítulo 5.

1 Otimização Multiobjetivo

A otimização escalar é a área da matemática que desenvolve métodos para encontrar os minimizadores (ou maximizadores) de uma função real em um conjunto viável (ver Definição 15 no Apêndice A)

Todo resultado obtido com minimizadores globais de funções pode ser utilizados para *maximizadores globais* de funções (ver Proposição 24 no Apêndice A).

A otimização multiobjetivo consiste em minimizar simultaneamente as funções componentes $f_j: \mathbb{R}^n \rightarrow \mathbb{R}, j \in \{1, \dots, m\}$, de uma função vetorial $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Portanto, a otimização escalar é o caso especial em que $m = 1$.

1.1 Ótimo de Pareto e ótimo fraco de Pareto

O contradomínio da função a ser minimizada é o conjunto \mathbb{R}^m , que não é totalmente ordenado para $m > 1$, e na maioria dos casos os objetivos são conflitantes. A seguinte ordem parcial é considerada.

Definição 1. A *ordem parcial* de \mathbb{R}^m é definida como

$$\begin{aligned} \forall u, v \in \mathbb{R}^m: \quad u \preceq v &\iff u_i \leq v_i, \quad \forall i \in \{1, \dots, m\} \\ u \prec v &\iff u_i < v_i, \quad \forall i \in \{1, \dots, m\}. \end{aligned}$$

Devido a limitação da ordem parcial, em vez de *minimizador* utiliza-se *ótimo de Pareto*, que consiste em um ponto $\tilde{x} \in C \subseteq \mathbb{R}^n$ tal que se existir $x \in C$ que reduz alguma função componente

$(\exists j \in \{1, \dots, m\}: f_j(x) < f_j(\tilde{x}))$ então ele aumenta alguma outra função componente $(\exists k \in \{1, \dots, m\}: f_k(x) > f_k(\tilde{x}))$ (GRÃÑA DRUMMOND; SVAITER, 2015).

Definição 2. Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Então $\tilde{x} \in C \subseteq \mathbb{R}^n$ é dito ser *ótimo de Pareto* ou *ponto não dominado* de f em C se

$$f(x) \preceq f(\tilde{x}) \implies f(x) = f(\tilde{x}), \quad \forall x \in C.$$

O conjunto de todos ótimos de Pareto é chamado *parede ótima* ou *fronteira de Pareto*.

A Figura 1 mostra um exemplo de ótimo de Pareto.

Graficamente, um ponto $x \in C \subseteq \mathbb{R}^n$ é um ótimo de Pareto em C quando não existe nenhum $y \in \mathbb{R}^m$, $y \preceq 0$, tal que $f(x) + y \in f(C)$, onde $f(C)$ é a imagem de f restrito a C .

Uma definição mais fraca de ótimo de Pareto é exigir apenas que não seja possível encontrar outro ponto que reduza todas funções componentes simultaneamente. Neste caso, chama-se *ótimo fraco de Pareto*, definido a seguir.

Definição 3. Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Então $\tilde{x} \in C \subseteq \mathbb{R}^n$ é *ótimo fraco de Pareto* de f em C se

$$\nexists x \in C: f(x) \prec f(\tilde{x}).$$

Em outras palavras, se existe um ponto $x \in C$ que reduz alguma função componente $(\exists j \in \{1, \dots, m\}: f_j(x) < f_j(\tilde{x}))$ então haverá outra função componente que foi aumentada ou permaneceu igual $(\exists k \in \{1, \dots, m\}: f_k(x) \geq f_k(\tilde{x}))$. A Figura 2 mostra um exemplo de ótimo fraco de Pareto.

Graficamente, um ponto $x \in \mathbb{R}^n$ é um ótimo fraco de Pareto em C quando $\nexists y \in \mathbb{R}^m$, $y \prec 0$ tal que $f(x) + y \in f(C)$.

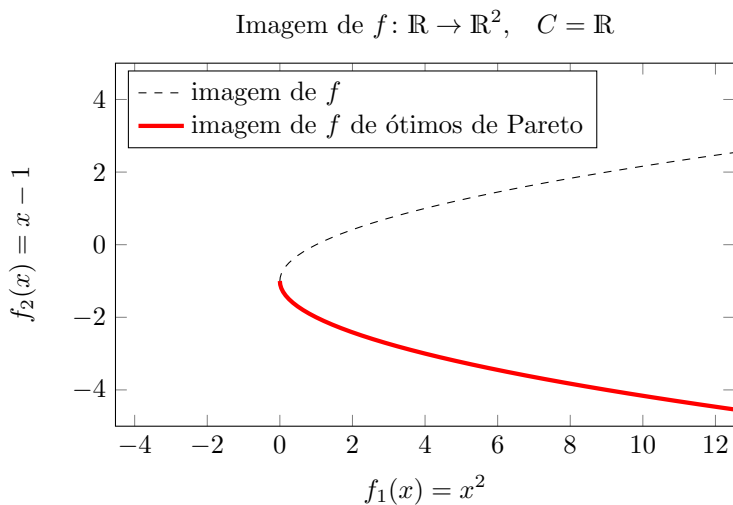


Figura 1 – O gráfico representa a imagem de f com destaque para a imagem de ótimos de Pareto. Por exemplo, quando $f_1(x) = x^2 = 1$ então $x = 1$ ou $x = -1$. Portanto, é possível $f_2(-1) = -2$ e $f_2(1) = 0$. Ambos pontos estão plotados no gráfico, mas o ponto $x = -1$ está classificado como ótimo de Pareto enquanto o $x = 1$ está classificado como comum.

Quando o domínio é unidimensional, o gráfico sempre formará uma curva. No caso de um domínio com mais dimensões, o gráfico pode se tornar uma região, como visto na Figura 3.

Note que a diferença entre ótimo de Pareto e ótimo fraco de Pareto é apenas a desigualdade ser ou não ser estrita. Por isso, segue diretamente da definição que todo ótimo de Pareto é um ótimo fraco de Pareto.

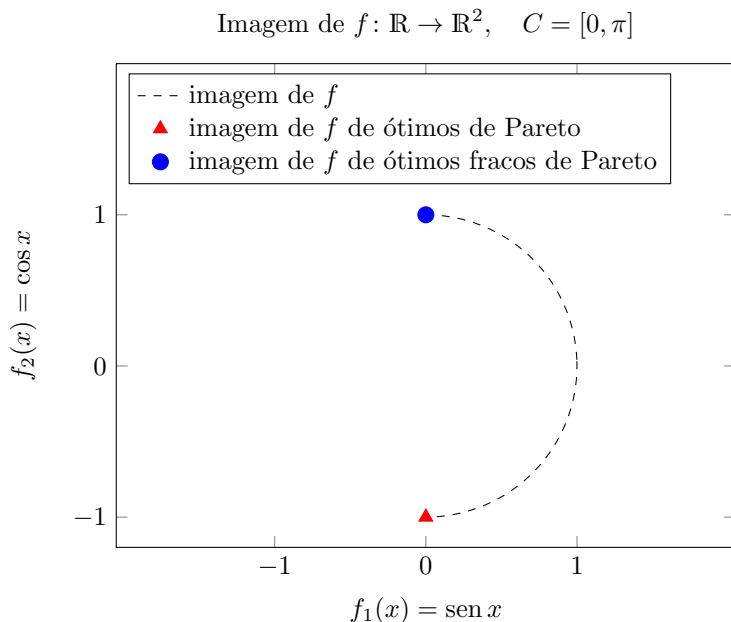


Figura 2 – O gráfico acima mostra a imagem de f restrita ao conjunto viável $C = [0, \pi]$, com destaque para a imagem do ótimo e ótimo fraco de Pareto. Quando $x = 0$ então $\cos 0 = 1$ e $\sin 0 = 0$, o que é um ótimo fraco de Pareto porque não é possível reduzir mais a função $\sin x$ em C . Mas é possível reduzir $\cos x$ sem aumentar $\sin x$: quando $x = \pi$, que é ótimo de Pareto.

1.2 Direção factível de descida e ponto crítico

Antes de explorar os métodos de otimização multiobjetivo é necessário explorar as características que os ótimos de Pareto possuem no caso de funções diferenciáveis em um conjunto viável $C \subseteq \mathbb{R}^n$ *convexo* (ver Definição 16 no Apêndice A).

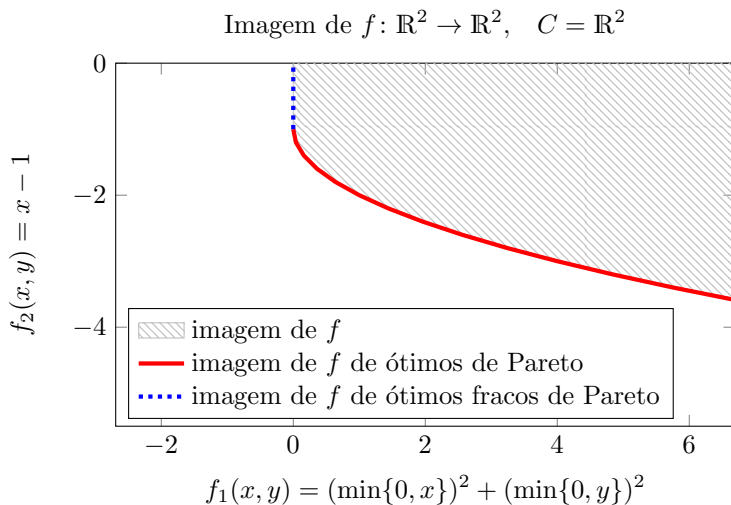


Figura 3 – O gráfico mostra a imagem de f com destaque para a imagem dos ótimos e ótimos fracos de Pareto. Quando $f_1(x, y) = 0$ não é possível reduzir mais f_1 e, portanto, todos pontos em que $f_1(x, y) = 0$ são ótimo fraco de Pareto, mas não necessariamente ótimo de Pareto.

Serão mostradas agora algumas notações que serão frequentemente usadas.

Notação 1. Os *vetores*, por exemplo $v \in \mathbb{R}^n$, são sempre considerados como vetores coluna, isto é, matrizes de 1 coluna e n linhas, e v^T indica o vetor transposto, isto é, um vetor linha, que é uma matriz de n colunas e 1 linha.

Será denotado por $\mathbf{1}^m \in \mathbb{R}^m$ o vetor em que todas componentes são iguais a 1.

Notação 2. A norma euclidiana, norma infinito e norma matricial (norma induzida) serão denotadas, respectivamente, por

$$\begin{aligned} \|x\| &:= \sqrt{x^T x}, & \forall x \in \mathbb{R}^n, \\ \|x\|_\infty &:= \max_{i \in \{1, \dots, n\}} |x_i|, & \forall x \in \mathbb{R}^n, \\ \|A\| &:= \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}, & \forall A \in \mathbb{R}^{m \times n}. \end{aligned}$$

Notação 3. As bolas abertas e fechadas de centro $x \in \mathbb{R}^n$ e raio $r \in \mathbb{R}$ serão denotadas, respectivamente, por

$$\begin{aligned} B(x, r) &:= \{y \in \mathbb{R}^n : \|x - y\| < r\}, \\ \overline{B(x, r)} &:= \{y \in \mathbb{R}^n : \|x - y\| \leq r\}. \end{aligned}$$

Uma característica que funções diferenciáveis possuem é a de serem aproximadamente lineares localmente. É fácil determinar uma direção em que uma função linear está decrescendo, basta tomar uma direção que faça um ângulo de mais de 90° com o gradiente. Essas direções são as direções de descida de uma função real. Em uma função vetorial, uma direção de descida é uma direção que é de descida para todas as funções componentes simultaneamente.

Para problemas com conjunto viável convexo, a factibilidade da direção de descida é importante. Por isso serão definidas direções de descida $v \in \mathbb{R}^n$ de $x \in C$ tal que $x + v \in C$, ou seja, factíveis.

Definição 4. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio e convexo, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C , $x \in C$ e $v \in \mathbb{R}^n$ tal que $x + v \in C$.

Então v é *direção factível de descida* de f em x se

$$\nabla f_j(x)^T v < 0, \quad \forall j \in \{1, \dots, m\},$$

onde $\nabla f_j(x)$ é o vetor gradiente da j -ésima função componente de f , isto é,

$$\nabla f_j(x)^T = \left(\frac{\partial f_j(x)}{\partial x_1}, \frac{\partial f_j(x)}{\partial x_2}, \dots, \frac{\partial f_j(x)}{\partial x_n} \right).$$

Se não houver direção factível de descida em x então x é dito *ponto crítico* ou *ponto estacionário* (FUKUDA; DRUMMOND, 2014).

Note que o produto escalar de dois vetores quaisquer $a, b \in \mathbb{R}^n$ pode ser interpretado como

$$a^T b = \|a\| \|b\| \cos \alpha,$$

onde α é o ângulo formado entre os vetores. Logo, $\nabla f_j(x)^T v < 0$ indica que o ângulo entre os vetores $\nabla f_j(x)$ e v é maior que 90° .

Uma forma equivalente de descrever uma direção factível de descida é através da matriz Jacobiana de f .

Notação 4. A matriz jacobiana de f é denotada $J_f: \mathbb{R}^n \rightarrow \mathbb{R}^{m \times n}$ e definida como

$$J_f(x) := \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}.$$

Então v é direção factível de descida de $f(x)$ se $x + v \in C$ e

$$J_f(x)v \prec 0.$$

Portanto, $x \in C$ é ponto crítico de f se

$$\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v \geq 0, \quad \forall v \in \mathbb{R}^n, x + v \in C.$$

Com estas definições, pode-se agora demonstrar que, se existe direção factível de descida $v \in \mathbb{R}^n$ em um ponto $x \in C$ então, próximo a este ponto, existe um ponto $x + tv \in C$ em que $f(x + tv) \prec f(x)$, descrito na proposição a seguir.

Proposição 1. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio e convexo, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C , e $v \in \mathbb{R}^n$ uma direção factível de descida para f em $x \in C$. Então:*

$$\exists \tilde{t} \in (0, 1): f(x + tv) \prec f(x), \quad \forall t \in (0, \tilde{t}).$$

Mais ainda, para qualquer $\sigma \in (0, 1)$:

$$\exists \tilde{t} \in (0, 1): f(x + tv) \prec f(x) + \sigma t J_f(x)v, \quad \forall t \in (0, \tilde{t}).$$

Um t que satisfaz a segunda inequação satisfaz a chamada condição de Armijo.

Demonstração. Por hipótese, f é diferenciável em um aberto contendo C .

Fixe $j \in \{1, \dots, m\}$. Então

$$\lim_{t \rightarrow 0^+} \frac{f_j(x + tv) - f_j(x)}{t} = \nabla f_j(x)^T v.$$

Seja $\sigma \in (0, 1)$. Como v é direção factível de descida então $\nabla f_j(x)^T v < 0$. Logo,

$$\lim_{t \rightarrow 0^+} \frac{f_j(x + tv) - f_j(x)}{t} < \nabla f_j(x)^T v < \sigma \nabla f_j(x)^T v < 0.$$

Então, pela definição de limite,

$$\exists \tilde{t}_j \in (0, 1): \frac{f_j(x + tv) - f_j(x)}{t} < \sigma \nabla f_j(x)^T v < 0, \quad \forall t \in (0, \tilde{t}_j).$$

Seja $\tilde{t} = \min_{j \in \{1, \dots, m\}} \tilde{t}_j$. Assim,

$$f(x + tv) - f(x) < \sigma t J_f(x) v < 0, \quad \forall t \in (0, \tilde{t}).$$

■

Pode-se concluir da Proposição 1 que, quando a função é diferenciável e o conjunto viável convexo, todo ótimo fraco de Pareto é um ponto crítico. É uma condição necessária porém não suficiente, isto é, nem todo ponto crítico é ótimo fraco de Pareto, como visto na Figura 4.

Um ponto crítico pode ser visto, portanto, como um ótimo fraco de Pareto localmente.

As funções convexas (ver Definição 18 no Apêndice B) estão relacionadas a ótimos de Pareto, como será visto na próxima proposição.

Proposição 2. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio e convexo, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C e $\tilde{x} \in C$ um ponto crítico. Então:*

- i. *f convexa em $C \implies \tilde{x}$ ótimo fraco de Pareto em C .*
- ii. *f estritamente convexa em $C \implies \tilde{x}$ ótimo de Pareto em C .*

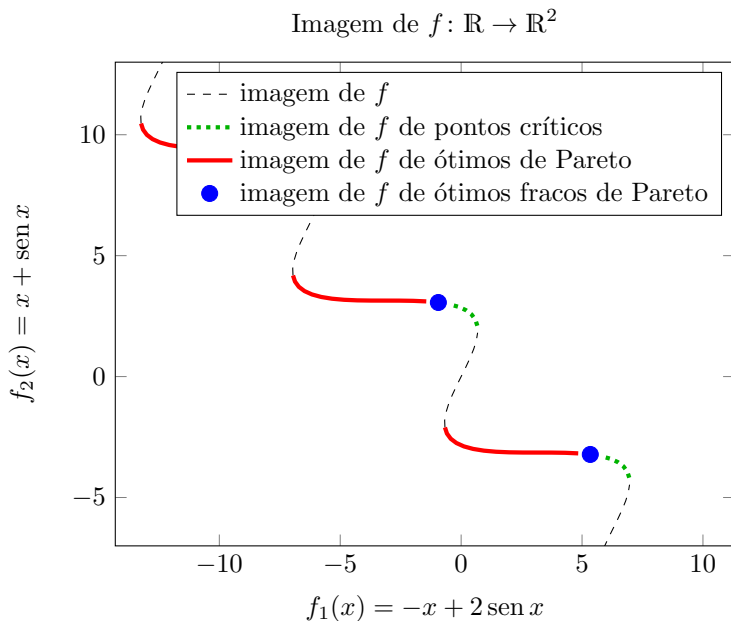


Figura 4 – O gráfico acima mostra a imagem de f com destaque para a imagem dos ótimos e ótimos fracos de Pareto, além da imagem dos pontos críticos.

Demonstração. i. Suponha que f é convexa e \tilde{x} não é ótimo fraco de Pareto.

Então $\exists x \in C: f(x) \prec f(\tilde{x})$.

Pela convexidade de f , para qualquer $\lambda \in (0, 1)$:

$$\begin{aligned} f(\lambda x + (1 - \lambda)\tilde{x}) &\preceq \lambda f(x) + (1 - \lambda)f(\tilde{x}) \preceq f(\tilde{x}) \\ \implies f(\tilde{x} + \lambda(x - \tilde{x})) &\prec f(\tilde{x}). \end{aligned}$$

Logo, $v = (x - \tilde{x})$ é uma direção factível de descida em \tilde{x} ,

pois $v + \tilde{x} \in C$, então \tilde{x} não é ponto crítico.

ii. Suponha que f é estritamente convexa.

Como foi mostrado, \tilde{x} é pelo menos ótimo fraco de Pareto.

Suponha que \tilde{x} não é ótimo de Pareto. Então

$$\exists x \in C: f(x) \preceq f(\tilde{x}) \wedge f(x) \neq f(\tilde{x}).$$

Pela convexidade estrita de f , se $\lambda \in (0, 1)$,

$$\begin{aligned} f(\lambda x + (1 - \lambda)\tilde{x}) &\prec \lambda f(x) + (1 - \lambda)f(\tilde{x}) \preceq f(\tilde{x}) \\ \implies f(\tilde{x} + \lambda(x - \tilde{x})) &\prec f(\tilde{x}). \end{aligned}$$

Logo, $v = (x - \tilde{x})$ é uma direção factível de descida em \tilde{x} , pois $v + \tilde{x} \in C$, então \tilde{x} não é ponto crítico, o que contraria o fato de ser ótimo fraco de Pareto. Portanto, \tilde{x} é ótimo de Pareto.



2 Convergência de métodos de descida multiobjetivo

Ser ponto crítico é condição necessária para ser um ótimo de Pareto, como foi visto na Proposição 1. Neste capítulo será definido um algoritmo que procura pontos críticos iterando por direções de descida em um conjunto viável convexo e fechado.

2.1 Método de descida

O protótipo geral de um método de descida está descrito no Algoritmo 1.

Algoritmo 1 (Direções de descida). Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C , $\sigma \in (0, 1)$, $x_0 \in C$ e $k = 0$.

[Passo 1] Se x_k é ponto crítico, PARE. Senão, encontre $v_k \in \mathbb{R}^n$ direção factível de descida de f em x_k .

[Passo 2] Encontre $t_k \in \mathbb{R}$, $t_k > 0$ tal que

$$f(x_k + t_k v_k) \preceq f(x_k) + \sigma t_k J_f(x_k) v_k.$$

[Passo 3] Faça $x_{k+1} = x_k + t_k v_k$

Faça $k = k + 1$.

Vá ao [Passo 1].

Note que o Algoritmo 1 está bem definido. Se x_k não é ponto crítico então existe direção factível de descida v_k . E, pela Proposição 1, deve existir um $t_k > 0$ que satisfaça a condição do [Passo 2].

Como foi dito, a condição exigida no [Passo 2] é chamada de *condição de Armijo* (NOCEDAL; WRIGHT, 1999). A ideia é que a função f não só reduza mas tenha uma redução suficiente, que é uma proporção da redução indicada pela derivada direcional. Apesar de ser possível $\sigma \in (0, 1)$, em termos práticos é tomado $0 < \sigma \leq \frac{1}{2}$.

No caso da otimização escalar, a sequência gerada no Algoritmo 1 converge para um ponto crítico se houverem duas condições extras para a direção factível de descida $v_k \in \mathbb{R}^n$, chamadas condição de proporcionalidade e condição do ângulo (MARTÍNEZ; SANTOS, 1995).

Para $f: \mathbb{R}^n \rightarrow \mathbb{R}$, a condição de proporcionalidade exige que, fixado $\beta > 0$,

$$\|v_k\| \geq \beta \|\nabla f(x_k)\|, \quad \forall k \in \mathbb{N}.$$

Já a condição do ângulo exige que, fixado $\theta \in (0, 1)$,

$$\nabla f(x_k)^T v_k < -\theta \|\nabla f(x_k)\| \|v_k\|, \quad \forall k \in \mathbb{N}.$$

Uma interpretação para a condição do ângulo é a exigência de um ângulo máximo permitido entre a direção factível de descida e o gradiente negativo. Esta condição não é possível no contexto multiobjetivo, pois pode existir um ponto x_k que possua direções factíveis de descida mas nenhuma satisfazendo a condição do ângulo para todos os gradientes ao mesmo tempo. Neste caso, o Algoritmo 1 não estaria bem definido. A Figura 5 ilustra um pouco melhor este problema.

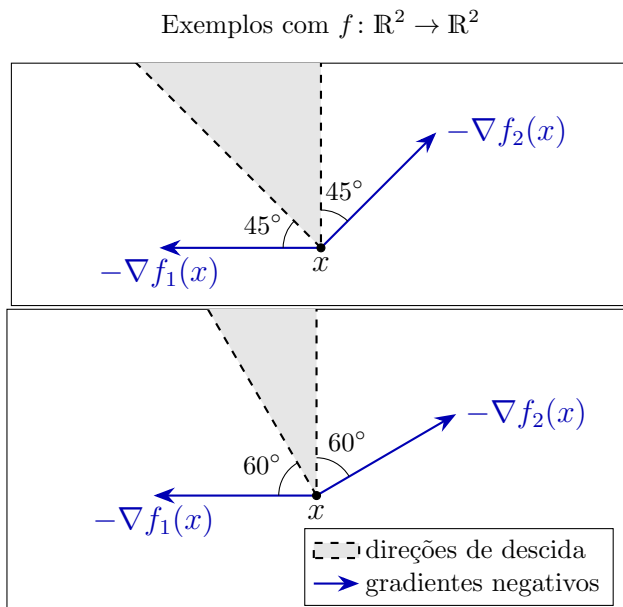


Figura 5 – Nestes exemplos, qualquer direção factível de descida $v \in \mathbb{R}^2$ forma um ângulo **mínimo** de 45° e 60° com cada gradiente negativo. A condição do ângulo define um ângulo **máximo** entre v e todos os gradientes negativos. Pode-se criar exemplos formando ângulos mínimos tão próximos de 90° quanto desejado. Portanto, a condição do ângulo, na otimização multiobjetivo, pode fazer com que nenhuma direção factível de descida possa ser escolhida em um ponto não crítico.

Sem a condição do ângulo, não é garantido que o Algoritmo 1 convirja para um ponto crítico na otimização escalar. E como a otimização escalar é um caso particular da otimização multiobjetivo, também não é garantida a convergência para um ponto crítico usando as hipóteses deste algoritmo apenas, como

é mostrado no Exemplo 1 (Apêndice C).

2.2 Método de descida com backtracking

A sequência $\{t_k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}$ do Algoritmo 1 pode ser obtida de forma simples com o método *backtracking*, que consiste em ir diminuindo o t_k a partir do 1 até satisfazer a condição exigida (GRÃÑA DRUMMOND; SVAITER, 2015). Assim, o Algoritmo 2, definido a seguir, é um caso particular do Algoritmo 1.

Algoritmo 2 (Direções de descida com backtracking). Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C , $\sigma \in (0, 1)$, $x_0 \in C$ e $k = 0$.

[Passo 1] Se x_k é ponto crítico, PARE. Senão, encontre $v_k \in \mathbb{R}^n$ direção factível de descida de f em x_k .

[Passo 2] Faça $t_k = 2^{-p_k}$, onde $p_k \in \mathbb{N}$ é o menor número a satisfazer

$$f(x_k + t_k v_k) \preceq f(x_k) + \sigma t_k J_f(x_k) v_k.$$

[Passo 3] Faça $x_{k+1} = x_k + t_k v_k$.

Faça $k = k + 1$.

Vá ao [Passo 1].

Pela Proposição 1, é garantido que o backtracking tem terminação finita, e portanto está bem definido.

Com o Algoritmo 2, basta encontrar uma forma de determinar uma direção factível de descida.

2.3 Aplicação regular de direções

Foi criada um tipo de aplicação chamada de *aplicação regular de direções* que define direções factíveis de descida, e que permite provar o Teorema 1, que será visto adiante, que afirma que todo ponto limite da sequência $\{x_k\}$ do Algoritmo 2 é ponto crítico, quando f for continuamente diferenciável.

A vantagem de analisar direções factíveis de descida de uma forma mais geral é facilitar a análise de outras direções factíveis de descida que serão definidas no Capítulo 4.

Definição 5. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C .

Então $\gamma: C \rightarrow \mathbb{R}^n$ é dita *aplicação regular de direções* para f se satisfizer as seguintes condições:

- i. $\gamma(x)$ é o vetor nulo quando x é crítico, e uma direção factível de descida quando x não é crítico.
- ii. Se $K \subseteq C$ é compacto então $\gamma(K)$ é limitado.
- iii. $\forall \{x_k\}_{k \in \mathbb{N}} \subseteq C$ pontos não críticos tal que $x_k \rightarrow x \in C$ não crítico, vale

$$\limsup_{k \rightarrow +\infty} J_f(x_k) \gamma(x_k) \prec 0.$$

Nas aplicações regulares de direções, a condição (i.) exige que $\gamma(C)$ contenha somente direções factíveis de descida e o vetor nulo, o que é algo natural em uma função que retorna direções factíveis de descida.

A condição (ii.) é usada para mostrar que, em pontos limites das sequências geradas pelo Algoritmo 2, os gradientes ficam

quase nulos, o que indica um ponto crítico.

A condição (iii.), juntamente com a condição (ii.), exige algo semelhante à condição do ângulo e condição da proporcionalidade juntas (discutido na Seção 2.1), pois para satisfazer essa condição, $\gamma(x_k) := v_k$ não pode ficar com a norma quase nula próximo a um ponto não crítico, e nem com $\gamma(x_k)$ quase ortogonal a nenhum dos gradientes, já que $\gamma(x_k)$ e os gradientes estão limitados sob compactos e dois vetores quase ortogonais com normas limitadas ficam com o produto interno quase nulo. Mas note que a condição (iii.) só se aplica para sequências convergindo para pontos não críticos, e portanto é uma condição um pouco mais fraca que a condição do ângulo.

Em (BERTSEKAS, 1999) é definido um conceito quase idêntico ao da Definição 5 para otimização escalar, que são as *sequências gradient-related*. A sequência $\{\gamma(x_k)\}$ é *gradient-related* se é uma sequência de direções factíveis de descida, limitada e que satisfaz a condição (iii.). Porém, supõe-se que f seja escalar ($f: \mathbb{R}^n \rightarrow \mathbb{R}$). Ainda em (BERTSEKAS, 1999), a sequência *gradient-related* é usada para demonstrar algo semelhante ao Teorema 1, mas novamente no caso escalar. Assim, podemos dizer que propomos uma generalização da condição *gradient-related* do caso escalar para o caso multiobjetivo.

As demonstrações de convergência apresentadas nas próximas seções foram inspiradas nas demonstrações de (FUKUDA; DRUMMOND, 2014) e (GRAÑA DRUMMOND; SVAITER, 2015).

2.3.1 Convergência em funções continuamente diferenciáveis

Nesta seção será mostrado que todo ponto limite da sequência $\{x_k\}$ gerada no Algoritmo 2 é um ponto crítico, se f for continu-

amente diferenciável, um ótimo fraco de Pareto, se f for convexa, e ótimo de Pareto, se f for estritamente convexa.

Um resultado interessante do Algoritmo 1 é que se a sequência $\{x_k\} \subseteq C$ gerada tem algum ponto limite então $\{f(x_k)\}$ converge, mesmo que $\{x_k\}$ não seja convergente.

Proposição 3. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C e $\{x_k\}_{k \in \mathbb{N}}$ uma sequência infinita gerada pelo Algoritmo 1.*

Se $\{x_k\}$ possui algum ponto limite então a sequência $\{f(x_k)\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^m$ é convergente.

Demonstração. Por construção, $f(x_{k+1}) \prec f(x_k), \forall k \in \mathbb{N}$.

Além disso, como $\{x_k\}$ tem ponto limite então existe uma subsequência convergente, ou seja, $\lim_{j \rightarrow +\infty} x_{k_j} \rightarrow \tilde{x} \in C$. Note que $\tilde{x} \in C$ pois C é fechado por hipótese.

Assim, como f é contínua, então $f(x_{k_j}) \rightarrow f(\tilde{x})$. Como $\{f(x_k)\}$ é decrescente, na ordem parcial do \mathbb{R}^m , e possui subsequência convergente então a sequência inteira converge para $f(\tilde{x})$, ou seja, $f(x_k) \rightarrow f(\tilde{x})$. ■

Outro resultado interessante é a garantia de existência de ponto limite na sequência gerada pelo Algoritmo 1 dada certas condições.

Proposição 4. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C , $\{x_k\}_{k \in \mathbb{N}}$ uma sequência infinita gerada pelo Algoritmo 1, $x_0 \in C$ o primeiro elemento de $\{x_k\}$ e $D \subseteq C$ tal que*

$$D := \{x \in C: f(x) \preceq f(x_0)\}.$$

Se D é limitado então $\{x_k\}$ possui ponto limite.

Demonstração. Como $\{f(x_k)\}$ é decrescente, na ordem parcial do \mathbb{R}^m , então $\{x_k\} \subseteq D$. Como D é limitado, pelo Teorema de Bolzano-Weierstrass, $\{x_k\}$ possui subsequência convergente. ■

A próxima proposição será para mostrar que o decréscimo exigido pela condição de Armijo (Algoritmo 1, [Passo 2]) forma uma sequência cuja série converge absolutamente.

Proposição 5. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C . Sejam $\{x_k\}_{k \in \mathbb{N}}$, $\{v_k\}_{k \in \mathbb{N}}$ e $\{t_k\}_{k \in \mathbb{N}}$ sequências infinitas geradas no Algoritmo 1 e suponha que $\{f(x_k)\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^m$ seja limitada. Então, para cada $j \in \{1, \dots, m\}$:*

$$\sum_{k=1}^{+\infty} t_k |\nabla f_j(x_k)^T v_k| < +\infty.$$

E portanto,

$$\lim_{k \rightarrow +\infty} t_k J_f(x_k) v_k = 0 \in \mathbb{R}^m.$$

Demonstração. Como a sequência $\{x_k\}$ foi gerada pelo Algoritmo 1, a seguinte desigualdade é satisfeita, para algum $\sigma \in (0, 1)$:

$$f_j(x_{k+1}) \leq f_j(x_k) + \sigma t_k \nabla f_j(x_k)^T v_k, \quad \forall k \in \mathbb{N}, \forall j \in \{1, \dots, m\}.$$

Fixe $j \in \{1, \dots, m\}$. Como $\{f_j(x_k)\}_{k \in \mathbb{N}}$ é decrescente por construção, então

$$\begin{aligned} -\sigma t_1 \nabla f_j(x_1)^T v_1 &\leq f_j(x_1) - f_j(x_2) \\ \implies -\sigma t_2 \nabla f_j(x_2)^T v_2 &\leq f_j(x_2) - f_j(x_3) \\ &\vdots \\ \implies \sum_{k=1}^N -\sigma t_k \nabla f_j(x_k)^T v_k &\leq f_j(x_1) - f_j(x_{N+1}) \\ \implies \sum_{k=1}^{+\infty} -\sigma t_k \nabla f_j(x_k)^T v_k &\leq \limsup_{N \rightarrow +\infty} f_j(x_1) - f_j(x_N). \end{aligned}$$

Como $\{f(x_k)\}$ é limitada por hipótese,

$$\sum_{k=1}^{+\infty} -\sigma t_k \nabla f_j(x_k)^T v_k < +\infty.$$

Por v_k ser direção factível de descida, $\nabla f_j(x_k)^T v_k < 0$, para todo $k \in \mathbb{N}$. Portanto,

$$\sum_{k=1}^{+\infty} t_k |\nabla f_j(x_k)^T v_k| < +\infty.$$

Se uma série converge absolutamente então a sequência que a gera converge para zero. Portanto, o limite de cada sequência, para cada função componente, é

$$\begin{aligned} \lim_{k \rightarrow +\infty} t_k \nabla f_i(x_k)^T v_k &= 0, \quad \forall i \in \{1, \dots, m\} \\ \implies \lim_{k \rightarrow +\infty} t_k J_f(x_k) v_k &= 0 \in \mathbb{R}^m. \end{aligned}$$

■

O lema abaixo afirma que, com as direções factíveis de descida geradas por uma aplicação regular de direções, $\exists \tilde{t} > 0$ tal que qualquer $t \in (0, \tilde{t})$ fixado satisfaz simultaneamente a desigualdade da condição de Armijo (Algoritmo 2, [Passo 2]) para todos pontos de uma vizinhança de um ponto não crítico.

Lema 1. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C , $\tilde{x} \in C$ um ponto não crítico, $\sigma \in (0, 1)$ e $\gamma: C \rightarrow \mathbb{R}^n$ aplicação regular de direções. Então*

$$\begin{aligned} \exists \tilde{t} \in (0, 1], \exists \delta > 0: \quad & f(x + tv) \preceq f(x) + \sigma t J_f(x) \gamma(x), \\ & \forall x \in B(\tilde{x}, \delta) \cap C, \quad \forall t \in (0, \tilde{t}). \end{aligned}$$

Demonstração. Pela continuidade de J_f em C e por \tilde{x} não ser ponto crítico, existe uma vizinhança de \tilde{x} em que todos pontos não são críticos. Logo,

$$\exists R > 0: \quad 0 \notin \gamma(V), \quad \text{onde } V := \overline{B(\tilde{x}, R)} \cap C.$$

Suponha que $\nexists y \in \mathbb{R}^m$, $y \prec 0$, tal que

$$J_f(x) \gamma(x) \prec y, \quad \forall x \in V. \quad (2.1)$$

Então, é possível construir uma sequência $\{x_k\}_{k \in \mathbb{N}} \subseteq V$ que converge para um ponto em V , pois V é compacto, tal que $\limsup_{k \rightarrow +\infty} J_f(x_k) \gamma(x_k) \not\prec 0$, mas isso contraria a condição (iii.) da definição de aplicação regular de direções.

Portanto, $\exists y \in \mathbb{R}^m$, $y \prec 0$ satisfazendo (2.1).

Pela condição (ii.) de aplicação regular de direções, $\gamma(V)$ é limitada. Então, defina

$$M = \sup_{x \in V} \|\gamma(x)\|$$

$$L = \max_{j \in \{1, \dots, m\}} y_j$$

$$\epsilon = \frac{1 - \sigma}{M}(-L).$$

Ainda por J_f ser contínua em C , segue que

$$\exists r \in (0, R):$$

$$\max_{j \in \{1, \dots, m\}} \|\nabla f_j(x) - \nabla f_j(z)\| < \epsilon, \quad \forall x, z \in B(\tilde{x}, r) \cap C.$$

$$\text{Tome } \delta = \frac{r}{2} \text{ e } \tilde{t} = \min \left\{ \frac{r}{2M}, 1 \right\}.$$

Dessa forma, para todo $x \in B(\tilde{x}, \delta) \cap C$ e $t \in (0, \tilde{t})$, vale:

$$\|x + t\gamma(x) - \tilde{x}\| \leq \|x - \tilde{x}\| + t\|\gamma(x)\| < \frac{r}{2} + \frac{r}{2M}M = r.$$

Logo, $x + t\gamma(x) \in B(\tilde{x}, r) \cap C \subseteq V$, $\forall t \in (0, \tilde{t})$, pois C é convexo, $x \in C$, $x + \gamma(x) \in C$, $t \in (0, \tilde{t})$ e $\tilde{t} \in (0, 1]$.

Pelo Teorema de Taylor com resto integral, pode-se obter a função $f_j(x + t\gamma(x))$, $j \in \{1, \dots, m\}$, como

$$\begin{aligned} f_j(x + t\gamma(x)) &= f_j(x) + \int_0^1 \nabla f_j(x + st\gamma(x))^T t\gamma(x) ds \\ &= f_j(x) + t\nabla f_j(x)^T \gamma(x) \\ &\quad + \int_0^1 (\nabla f_j(x + st\gamma(x)) - \nabla f_j(x))^T t\gamma(x) ds. \end{aligned}$$

Usando a desigualdade de Cauchy-Schwarz na integral:

$$\begin{aligned} f_j(x + t\gamma(x)) &\leq f_j(x) + t\nabla f_j(x)^T \gamma(x) \\ &\quad + t\|\gamma(x)\| \int_0^1 \|\nabla f_j(x + st\gamma(x)) - \nabla f_j(x)\| ds \\ &\leq f_j(x) + t\nabla f_j(x)^T \gamma(x) + t\|\gamma(x)\| \int_0^1 \epsilon ds \\ &\leq f_j(x) + t\nabla f_j(x)^T \gamma(x) + tM\epsilon, \end{aligned}$$

e portanto,

$$f_j(x + t\gamma(x)) \leq f_j(x) + t \left(\nabla f_j(x)^T \gamma(x) + (1 - \sigma)(-L) \right). \quad (2.2)$$

Note que, de (2.1),

$$\begin{aligned} \nabla f_j(x)^T \gamma(x) &\leq y_j \leq L \\ \implies (1 - \sigma) \nabla f_j(x)^T \gamma(x) &\leq (1 - \sigma)L \\ \implies \nabla f_j(x)^T \gamma(x) + (1 - \sigma)(-L) &\leq \sigma \nabla f_j(x)^T \gamma(x). \end{aligned}$$

Utilizando este resultado em (2.2),

$$f_j(x + t\gamma(x)) \leq f_j(x) + t\sigma \nabla f_j(x)^T \gamma(x).$$

Como j foi arbitrário,

$$f(x + t\gamma(x)) \leq f(x) + \sigma t J_f(x) \gamma(x), \forall x \in B(\tilde{x}, \delta) \cap C, \forall t \in (0, \tilde{t}).$$

■

Teorema 1. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C . Sejam $\{x_k\}_{k \in \mathbb{N}}$, $\{v_k\}_{k \in \mathbb{N}}$ e $\{t_k\}_{k \in \mathbb{N}}$ sequências infinitas geradas no Algoritmo 2.*

Suponha que $\exists \tilde{x} \in C$ ponto limite de $\{x_k\}$ e que $\exists \gamma: C \rightarrow \mathbb{R}^n$ aplicação regular de direções tal que $v_k = \gamma(x_k)$, $\forall k \in \mathbb{N}$.

Então \tilde{x} é crítico. Se f for convexa em C , \tilde{x} é ótimo fraco de Pareto, e se f for estritamente convexa em C , \tilde{x} é ótimo de Pareto.

Demonstração. Como \tilde{x} é ponto limite de $\{x_k\}$, $\exists \{k_j\}_{j \in \mathbb{N}} \subseteq \mathbb{N}$ crescente onde $\{x_{k_j}\}_{j \in \mathbb{N}}$ converge para \tilde{x} .

Pela Proposição 3, como $\{x_k\}$ possui ponto limite, $\{f(x_k)\}$ é convergente e, portanto, limitada. Então vale a Proposição 5:

$$\lim_{k \rightarrow +\infty} t_k J_f(x_k) v_k = 0,$$

e restringida à subsequência,

$$\lim_{j \rightarrow +\infty} t_{k_j} J_f(x_{k_j}) v_{k_j} = 0. \quad (2.3)$$

Suponha que \tilde{x} não seja ponto crítico.

Então, pelo Lema 1, $\exists \tilde{t} \in (0, 1]$ e $\exists \delta > 0$ tal que se $x \in B(\tilde{x}, \delta) \cap C$ e $t \in (0, \tilde{t})$, a condição do [Passo 2] é satisfeita.

Como $x_{k_j} \rightarrow \tilde{x}$, $\exists N \in \mathbb{N}$ tal que, $\forall j > N$, $x_{k_j} \in B(\tilde{x}, \delta) \cap C$.

A variável t_{k_j} é definida como $2^{-p_{k_j}}$ onde $p_{k_j} \in \mathbb{N}$ é o menor número que faz com que t_{k_j} satisfaça o [Passo 2]. Então existe um $p \in \mathbb{N}$ onde $2^{-p} \in (0, \tilde{t})$ e, $\forall j > N$, $p > p_{k_j}$, ou seja, $2^{-p} \leq t_{k_j}$. Assim, $\{t_{k_j}\}$ possui uma cota inferior maior que zero para $j > N$. Logo, (2.3) pode ser reduzida a

$$\lim_{j \rightarrow +\infty} J_f(x_{k_j}) v_{k_j} = 0. \quad (2.4)$$

Como $x_{k_j} \rightarrow \tilde{x}$ e J_f é contínua em C então pela condição (iii.) da aplicação regular de direções,

$$\limsup_{j \rightarrow +\infty} J_f(x_{k_j}) v_{k_j} < 0,$$

que contraria (2.4). Portanto, \tilde{x} é um ponto crítico.

Se f for convexo em C , pela Proposição 2, \tilde{x} é ótimo fraco de Pareto, e for estritamente convexa em C , \tilde{x} é ótimo de Pareto. ■

2.3.2 Convergência em funções convexas diferenciáveis

Nesta seção será mostrado que toda sequência $\{x_k\}$ gerada no Algoritmo 2 converge para um ótimo fraco de Pareto se f for convexa e satisfizer a Hipótese 1, descrita a seguir.

Supor que todas funções componentes de f sejam limitadas inferiormente é muito restritivo. A Hipótese 1 é um pouco mais fraca, e supõe que toda sequência decrescente, na ordem parcial do \mathbb{R}^m , seja limitada inferiormente por um ponto da imagem do conjunto viável. Assim, as funções componentes podem ser ilimitadas inferiormente, mas não em uma sequência de ordem parcial do \mathbb{R}^m , que exige que todas reduzam ao mesmo tempo.

Hipótese 1. Sejam $C \subseteq \mathbb{R}^n$ não vazio e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Então $\forall \{f(x_k)\}_{k \in \mathbb{N}} \subseteq f(C)$ decrescente, na ordem parcial do \mathbb{R}^m :

$$\exists x \in C: \quad f(x) \preceq f(x_k), \quad \forall k \in \mathbb{N}.$$

Será definido o conceito de funções *coercivas*.

Definição 6. Sejam $C \subseteq \mathbb{R}^n$ não vazio e $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Então f é dita *coerciva* em C se, $\forall \{x_k\}_{k \in \mathbb{N}} \subseteq C$ tal que $\|x_k\| \rightarrow +\infty$,

$$\lim_{k \rightarrow +\infty} f(x_k) = +\infty.$$

Se C é limitado, f é considerada coerciva por vacuidade.

As funções contínuas coercivas possuem uma propriedade importante: elas sempre possuem um minimizador global em conjuntos convexos fechados e não vazios.

Proposição 6. Sejam $C \subseteq \mathbb{R}^n$ convexo, fechado e não vazio, e $f: \mathbb{R}^n \rightarrow \mathbb{R}$ contínua e coerciva em C .

Então existe um minimizador global de f em C .

Demonstração. Existem dois casos.

Caso C seja limitado: então, como C é fechado em \mathbb{R}^n , C é compacto. Pela continuidade de f em C , $f(C)$ é compacto, e portanto, existe minimizador global de f em C .

Caso C seja ilimitado: seja $x_0 \in C$.

Como f é coerciva, $\exists r > 0$ tal que

$$f(x) \geq f(x_0), \quad \forall x \in C \setminus \overline{B(x_0, r)}.$$

Pelo caso anterior, $\exists \tilde{x} \in \overline{B(x_0, r)} \cap C$ minimizador global de f em $\overline{B(x_0, r)} \cap C$.

Seja $x \in C \setminus \overline{B(x_0, r)}$. Então $f(x) \geq f(x_0) \geq f(\tilde{x})$.

Portanto, \tilde{x} é minimizador global de f em C .

■

Outra proposição importante é a de que funções estritamente convexas possuem, no máximo, um minimizador global em conjuntos convexos.

Proposição 7. *Sejam $C \subseteq \mathbb{R}^n$ convexo e não vazio, e $f: \mathbb{R}^n \rightarrow \mathbb{R}$ estritamente convexa em C .*

Então f admite no máximo um minimizador global em C , isto é, se \tilde{x} é minimizador global de f em C então

$$f(x) = f(\tilde{x}) \implies \tilde{x} = x, \quad \forall x \in C.$$

Demonstração. Seja $\tilde{x}, x \in C$ e $x \neq \tilde{x}$. Suponha que $f(x) = f(\tilde{x})$.

Pela convexidade estrita, para qualquer $\lambda \in (0, 1)$,

$$f(\lambda \tilde{x} + (1 - \lambda)x) < \lambda f(\tilde{x}) + (1 - \lambda)f(x) = f(\tilde{x}) = f(x).$$

Logo, nem \tilde{x} nem x são minimizadores globais de f em C . ■

Outro conceito que será usado é a de projeções em conjuntos convexos fechados, e suas propriedades.

Proposição 8. *Sejam $C \subseteq \mathbb{R}^n$ convexo, fechado e não vazio, e $\Pi_C: \mathbb{R}^n \rightarrow C$ tal que*

$$\Pi_C(x) := \underset{y \in C}{\operatorname{argmin}} \|x - y\|^2,$$

onde $\|x - y\|^2 = (x - y)^T(x - y)$.

Então Π_C é chamada projeção em C e está bem definida.

Além disso, $\forall x \in \mathbb{R}^n$, $z = \Pi_C(x)$ se e somente se

$$(y - z)^T(x - z) \leq 0, \quad \forall y \in C.$$

Demonstração. Será mostrado que Π_C está bem definida. Note que $\Pi_C(x)$ é o minimizador da função $d: C \rightarrow \mathbb{R}$,

$$d(y) := (x - y)^T(x - y) = \|x - y\|^2.$$

Note que d é estritamente convexa pois, $\forall y, z \in \mathbb{R}^n$, $y \neq z$ e $\forall \lambda \in (0, 1)$:

$$\begin{aligned} d(y + \lambda(z - y)) &= \|x - y - \lambda(z - y)\|^2 \\ &= \|x - y\|^2 + \lambda^2\|z - y\|^2 - 2\lambda(x - y)^T(z - y) \\ &< \|x - y\|^2 + \lambda\|z - x + x - y\|^2 - 2\lambda(x - y)^T(z - y) \\ &= d(y) + \lambda d(z) + \lambda d(y) - 2\lambda(x - y)^T((x - z) + (z - y)) \\ &= d(y) + \lambda d(z) + \lambda d(y) - 2\lambda d(y) \\ &= d(y) + \lambda(d(z) - d(y)). \end{aligned}$$

Além disso, d é trivialmente coerciva, portanto pelas Proposições 6 e 7, d possui um único minimizador global. Logo, Π_C está bem definida.

Fixe $x \in \mathbb{R}^n$ e seja $z = \Pi_C(x)$. Tome $y \in C$ qualquer.

Considere uma função $g: [0, 1] \rightarrow C$ definida como:

$$g(\lambda) = z + \lambda(y - z).$$

Então $g([0, 1]) \subseteq C$. Considere agora a função $h: [0, 1] \rightarrow \mathbb{R}$

$$\begin{aligned} h(\lambda) &= \|x - g(\lambda)\|^2 \\ &= \|(x - z) - \lambda(y - z)\|^2 \\ &= \|x - z\|^2 + \lambda^2\|y - z\|^2 - 2\lambda(x - z)^T(y - z). \end{aligned}$$

Como $z = \Pi_C(x) = g(0)$ e $\Pi_C(x)$ minimiza $d(y) = \|x - y\|^2$ então em particular, minimiza h , pois h é só uma restrição de d . Assim, $h(0)$ é o mínimo de h . Logo, para $\lambda \in (0, 1)$,

$$\begin{aligned} 0 &\leq \frac{h(\lambda) - h(0)}{\lambda} \\ \implies 0 &\leq \lim_{\lambda \rightarrow 0^+} \frac{h(\lambda) - h(0)}{\lambda} = h'(0) = -2(x - z)^T(y - z) \\ \implies (x - z)^T(y - z) &\leq 0. \end{aligned}$$

Para a recíproca, seja $z \in C$ e $x \in C$ tais que

$$(y - z)^T(x - z) \leq 0, \quad \forall y \in C.$$

Logo,

$$\begin{aligned} \|y - x\|^2 &= \|y - z + z - x\|^2 \\ &= \|y - z\|^2 + \|z - x\|^2 + 2(y - z)^T(z - x) \\ &\geq \|z - x\|^2 - 2(y - z)^T(x - z) \geq \|z - x\|^2 \\ \implies \|y - x\|^2 &\geq \|z - x\|^2, \quad \forall y \in C \\ \implies z &= \Pi_C(x). \end{aligned}$$

■

Para garantir a convergência das sequências geradas, será usado o conceito de *sequência monotônica Fejér*. Intuitivamente, uma sequência $\{x_k\}_{k \in \mathbb{N}}$ é monotônica Fejér a um conjunto D quando, para cada $x \in D$, a sequência da distância de x_k até x é decrescente. (COMBETTES, 2001)

Definição 7. A sequência $\{x_k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ é dita *monotônica Fejér* a $D \subseteq \mathbb{R}^n$ se $\forall x \in D$:

$$\|x_{k+1} - x\| \leq \|x_k - x\|, \quad \forall k \in \mathbb{N}.$$

Admitindo uma certa margem de erro nesta desigualdade obtém-se a chamada sequência *monotônica quasi-Fejér*. Há vários tipos de sequência monotônica quasi-Fejér, mas o que será usado é o *tipo III* da referência (COMBETTES, 2001), definido a seguir.

Definição 8. A sequência $\{x_k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ é dita *monotônica quasi-Fejér* a $D \subseteq \mathbb{R}^n$ se $\forall x \in D$, $\exists \{\epsilon_k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}$ não negativa tal que:

$$\sum_{i=0}^{+\infty} \epsilon_i < +\infty \wedge \|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 + \epsilon_k, \quad \forall k \in \mathbb{N}.$$

Se $\forall x \in D$ a sequência $\{\epsilon_k\}_{k \in \mathbb{N}}$ associada for identicamente nula, obtém-se uma sequência monotônica Fejér a D .

As sequências monotônicas quasi-Fejér a $D \neq \emptyset$ possuem algumas propriedades interessantes que são usadas na análise da convergência dos algoritmos apresentados.

Lema 2. *Seja $\{x_k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ sequência monotônica quasi-Fejér a $D \subseteq \mathbb{R}^n$ não vazio.*

Então $\{x_k\}$ é limitada.

Se $\exists x \in D$ ponto limite de $\{x_k\}$ então $x_k \rightarrow x$.

Demonstração. Essa demonstração é baseada na demonstração feita em ([GRAÑA DRUMMOND; SVAITER, 2015](#)).

Seja $x \in D$ e $\{\epsilon_k\}$ da Definição 8 associado a x .

Então, para um $k \in \mathbb{N}$ fixado:

$$\begin{aligned}
 \|x_{k+1} - x\|^2 &\leq \|x_k - x\|^2 + \epsilon_k \\
 \implies \|x_{k+2} - x\|^2 &\leq \|x_{k+1} - x\|^2 + \epsilon_{k+1} \\
 &\leq \|x_k - x\|^2 + \epsilon_k + \epsilon_{k+1} \\
 &\vdots \\
 \implies \|x_{k+j+1} - x\|^2 &\leq \|x_k - x\|^2 + \sum_{i=k}^{k+j} \epsilon_i \\
 \implies \limsup_{j \rightarrow +\infty} \|x_{k+j+1} - x\|^2 &\leq \|x_k - x\|^2 + \sum_{i=k}^{+\infty} \epsilon_i < +\infty,
 \end{aligned}$$

portanto,

$$\begin{aligned}
 \exists L > 0: \quad \limsup_{k \rightarrow +\infty} \|x_k - x\|^2 &\leq L^2 \\
 \implies \limsup_{k \rightarrow +\infty} \|x_k - x\| &\leq L.
 \end{aligned} \tag{2.5}$$

Pela desigualdade triangular,

$$\begin{aligned}
 \|x_k\| - \|x\| &\leq \|x_k - x\| \\
 \implies \limsup_{k \rightarrow +\infty} \|x_k\| - \|x\| &\leq \limsup_{k \rightarrow +\infty} \|x_k - x\| \leq L \\
 \implies \limsup_{k \rightarrow +\infty} \|x_k\| &< +\infty,
 \end{aligned}$$

portanto, $\{x_k\}$ é uma sequência limitada.

Suponha agora $\exists \tilde{x} \in D$ ponto limite de $\{x_k\}$. Seja $\{\tilde{\epsilon}_k\}$ associado a \tilde{x} . Será mostrado que $x_k \rightarrow \tilde{x}$. Fixe $\epsilon > 0$.

Como $0 \leq \sum_{k=0}^{+\infty} \tilde{\epsilon}_k < +\infty$ então

$$\exists k_0 \in \mathbb{N}: \sum_{k=k_0}^{+\infty} \tilde{\epsilon}_k < \frac{\epsilon}{2}.$$

Além disso, $\exists k_1 > k_0$ tal que $\|x_{k_1} - \tilde{x}\|^2 < \frac{\epsilon}{2}$, pois \tilde{x} é um ponto limite de $\{x_k\}$.

Como $\{x_k\}$ é sequência monotônica quasi-Fejér a D e $\tilde{x} \in D$,

$$\|x_k - \tilde{x}\|^2 \leq \|x_{k_1} - \tilde{x}\|^2 + \sum_{i=k_1}^{k-1} \tilde{\epsilon}_i < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \quad \forall k > k_1.$$

Logo, $x_k \rightarrow \tilde{x}$. ■

Uma definição um pouco mais forte que a de aplicação regular de direções será dada.

Definição 9. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C .

Então $\gamma: C \rightarrow \mathbb{R}^n$ é dita *aplicação fortemente regular de direções* para f se for aplicação regular de direções de f e ainda, quando f for convexa, $\forall x \in C, \exists w \in [0, 1]^m$:

$$\gamma(x) = \Pi_C(x - J_f(x)^T w) - x,$$

em que Π_C é a projecção em C , da Proposição 8.

Nas aplicações de direções fortemente regulares, a direcção factível de descida é uma combinação cônica limitada dos gradientes negativos projetada no conjunto viável trasladado para

origem. Isso vai garantir adiante que a sequência gerada pelo Algoritmo 2 seja convergente, dada certas condições sobre a função f .

O lema abaixo auxiliará na ligação entre sequências monotônicas quasi-Fejér e as sequências geradas no Algoritmo 2.

Lema 3. *Sejam $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável e convexa em um aberto contendo C . Suponha valer a Hipótese 1. Sejam $\{x_k\}_{k \in \mathbb{N}}$, $\{v_k\}_{k \in \mathbb{N}}$ e $\{t_k\}_{k \in \mathbb{N}}$ sequências infinitas geradas no Algoritmo 2.*

Suponha que $\exists \gamma: C \rightarrow \mathbb{R}^n$ aplicação fortemente regular de direções tal que $v_k = \gamma(x_k)$, $\forall k \in \mathbb{N}$.

Seja $x \in C$ tal que $f(x) \preceq f(x_k)$, $\forall k \in \mathbb{N}$ da Hipótese 1. Então

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 + 2t_k \sum_{j=1}^m |\nabla f_j(x_k)^T v_k|.$$

Demonstração. Como f é convexa e diferenciável em um aberto contendo C , pela Proposição 25 (no Apêndice B),

$$J_f(x_k)v \preceq f(x_k + v) - f(x_k), \quad \forall v \in (C - x_k). \quad (2.6)$$

onde $(C - x_k) := \{v \in \mathbb{R}^n : x_k + v \in C\}$.

Seja $v = x - x_k$. Então $v \in (C - x_k)$.

Por γ ser aplicação fortemente regular de direções e f ser convexa, $\exists w_k \in [0, 1]^m$ tal que

$$v_k = \gamma(x_k) = \Pi_C(x_k - J_f(x_k)^T w_k) - x_k.$$

Então, pela Proposição 8,

$$(x_k - J_f(x_k)^T w_k - (v_k + x_k))^T (x_k + v - (v_k + x_k)) \leq 0$$

$$\begin{aligned}
&\implies (-J_f(x_k)^T w_k - v_k)^T (v - v_k) \leq 0 \\
&\implies -w_k^T J_f(x_k) v + w_k^T J_f(x_k) v_k - v_k^T v + \|v_k\|^2 \leq 0 \\
&\implies w_k^T J_f(x_k) v_k - v_k^T v + \|v_k\|^2 \leq w_k^T J_f(x_k) v. \tag{2.7}
\end{aligned}$$

Por (2.6) e por $w_k \succeq 0$

$$\begin{aligned}
&J_f(x_k) v \preceq f(x_k + v) - f(x_k) = f(x) - f(x_k) \preceq 0 \\
&\implies w_k^T J_f(x_k) v \leq 0. \tag{2.8}
\end{aligned}$$

Então, de (2.7) e (2.8),

$$\begin{aligned}
&-v_k^T v \leq -w_k^T J_f(x_k) v_k - \|v_k\|^2 \\
&\implies \|v_k\|^2 - v_k^T v \leq |w_k^T J_f(x_k) v_k|. \tag{2.9}
\end{aligned}$$

Pelo Algoritmo 2, e usando que $x_k + v = x$ e $t_k \in (0, 1]$,

$$\begin{aligned}
\|x_{k+1} - x\|^2 &= \|x_k + t_k v_k - x\|^2 \\
&= \|x_k - x\|^2 + t_k^2 \|v_k\|^2 - 2t_k v_k^T (x - x_k) \\
&\leq \|x_k - x\|^2 + 2t_k \|v_k\|^2 - 2t_k v_k^T (x - x_k) \\
&= \|v\|^2 + 2t_k (\|v_k\|^2 - v_k^T v).
\end{aligned}$$

Usando o resultado de (2.9) na equação acima, tem-se

$$\begin{aligned}
\|x_{k+1} - x\|^2 &\leq \|v\|^2 + 2t_k |w_k^T J_f(x_k) v_k| \\
&= \|x_k - x\|^2 + 2t_k |w_k^T J_f(x_k) v_k|.
\end{aligned}$$

Como $w_k \in [0, 1]^m$,

$$\begin{aligned}
|w_k^T J_f(x_k) v_k| &= \left| \sum_{j=1}^m (w_k)_j \nabla f_j(x_k)^T v_k \right| \\
&\leq \sum_{j=1}^m |(w_k)_j| |\nabla f_j(x_k)^T v_k| \leq \sum_{j=1}^m |\nabla f_j(x_k)^T v_k|.
\end{aligned}$$

Assim,

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 + 2t_k \sum_{j=1}^m |\nabla f_j(x_k)^T v_k|.$$

■

Teorema 2. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável e convexa em um aberto contendo C . Suponha também valer a Hipótese 1.*

Sejam $\{x_k\}_{k \in \mathbb{N}}$ e $\{v_k\}_{k \in \mathbb{N}}$ sequências infinitas geradas no Algoritmo 2.

Suponha que $\exists \gamma: C \rightarrow \mathbb{R}^n$ aplicação fortemente regular de direções tal que $v_k = \gamma(x_k)$, $\forall k \in \mathbb{N}$.

Então, $x_k \rightarrow \tilde{x} \in C$ ótimo fraco de Pareto.

Se f for estritamente convexa em C , \tilde{x} é ótimo de Pareto.

Demonstração. Defina o conjunto

$$D := \{x \in \mathbb{R}^n: f(x) \preceq f(x_k), \quad \forall k \in \mathbb{N}\}.$$

A sequência $\{f(x_k)\}_{k \in \mathbb{N}}$ é decrescente, na ordem parcial do \mathbb{R}^m , e f satisfaz a Hipótese 1. Por isso, $D \neq \emptyset$. Será mostrado agora que $\{x_k\}$ é uma sequência monotônica quasi-Fejér a D .

Seja $x \in D$. Como γ é aplicação fortemente regular de direções, pelo Lema 3,

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 + 2t_k \sum_{j=1}^m |\nabla f_j(x_k)^T v_k|.$$

Tome $\epsilon_k = 2t_k \sum_{j=1}^m |\nabla f_j(x_k)^T v_k|$. Como vale a Hipótese 1 então $\{f(x_k)\}$ é limitada. Logo, vale a Proposição 5:

$$\sum_{k=1}^{+\infty} t_k |\nabla f_j(x_k)^T v_k| < +\infty, \quad \forall j \in \{1, \dots, m\}.$$

Como a série acima converge absolutamente,

$$\begin{aligned} \sum_{k=1}^{+\infty} \epsilon_k &= \sum_{k=1}^{+\infty} \left(2t_k \sum_{j=1}^m |\nabla f_j(x_k)^T v_k| \right) \\ &= 2 \sum_{j=1}^m \sum_{k=1}^{+\infty} t_k |\nabla f_j(x_k)^T v_k| < +\infty. \end{aligned}$$

Portanto, a sequência $\{x_k\}$ é monotônica quasi-Fejér a D não vazio. Então pelo Lema 2, $\{x_k\}$ é limitada, o que implica a existência de $\tilde{x} \in \mathbb{R}^n$ ponto limite de $\{x_k\}$. E como $\tilde{x} \in D$, ainda pelo Lema 2, $x_k \rightarrow \tilde{x}$.

Pela Proposição 26 (no Apêndice B), f é continuamente diferenciável em um aberto contendo C . Então, pelo Teorema 1, \tilde{x} é ponto crítico.

Pela Proposição 2, \tilde{x} é ótimo fraco de Pareto e, se f for estritamente convexa em C , \tilde{x} é ótimo de Pareto.

■

2.3.3 Algoritmo do método de descida com aplicação regular de direções

O Algoritmo 3, mostrado a seguir, é um caso particular do Algoritmo 2, onde a direção factível de descida é determinada por uma aplicação regular de direções. O critério de parada usado, $J_f(x_k)\gamma(x_k) \not\approx 0$, em muitos casos só é satisfeito quando

$k \rightarrow +\infty$. Como é necessário que o algoritmo termine, o critério de parada utilizado na prática é $J_f(x_k)\gamma(x_k) \not\leq -\epsilon \mathbb{1}^n$, para algum $\epsilon > 0$ fixado.

Algoritmo 3 (Aplicação regular de direções). Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C .

Sejam $\gamma: C \rightarrow \mathbb{R}^n$ uma aplicação regular de direções (ver Definição 5), $\sigma \in (0, 1)$, $x_0 \in C$ e $k = 0$.

[Passo 1] Se $J_f(x_k)\gamma(x_k) \not\leq 0$, PARE.

[Passo 2] Faça $t_k = 2^{-p_k}$, onde $p_k \in \mathbb{N}$ é o menor número a satisfazer

$$f(x_k + t_k \gamma(x_k)) \preceq f(x_k) + \sigma t_k J_f(x_k) \gamma(x_k).$$

[Passo 3] Faça $x_{k+1} = x_k + t_k \gamma(x_k)$.

Faça $k = k + 1$.

Vá ao [Passo 1].

3 Direção de Cauchy multiobjetivo

Com os conceitos vistos, será possível agora definir e analisar a *direção de Cauchy multiobjetivo* (GRAÑA DRUMMOND; SVAITER, 2015; FUKUDA; DRUMMOND, 2014).

Considerando $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ função diferenciável, $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado o conjunto viável, $x \in C$ e $v \in \mathbb{R}^n$ tal que $x + v \in C$.

Então pelo Teorema de Taylor vale a igualdade

$$f(x + v) = f(x) + J_f(x)v + h(x + v)\|v\|^2,$$

onde $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ é uma função em $h(x + v) \rightarrow 0$ quando $v \rightarrow 0$.

Portanto, pode-se estimar $f(x + v)$ por

$$f(x + v) \approx f(x) + J_f(x)v,$$

com um erro tão pequeno quanto desejado para v de norma suficientemente pequena.

Considerando esta aproximação, pode-se afirmar que o problema de encontrar v que mais reduz $f(x)$ é o problema de minimizar, na ordem parcial do \mathbb{R}^m , a função $J_f(x)v$. Para transformar este problema de otimização multiobjetivo em otimização escalar, será minimizado $\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$. Considere então o problema:

$$\underset{\substack{v \in \mathbb{R}^n \\ x+v \in C}}{\operatorname{argmin}} \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v + \frac{1}{2} \|v\|^2. \quad (3.1)$$

O termo $\frac{1}{2}\|v\|^2$ serve para limitar a norma de v .

Note que para o caso escalar $f: \mathbb{R}^n \rightarrow \mathbb{R}$ com $C = \mathbb{R}^n$, o problema (3.1) atinge o mínimo quando $v = -\nabla f(x)$, que é justamente a direção de Cauchy na otimização escalar. No caso multiobjetivo, o problema pode ser resolvido de forma iterativa, e os detalhes serão mostrados mais a frente.

Definição 10. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C e $x \in C$.

Então $\tilde{v} \in \mathbb{R}^n$ é uma *direção de Cauchy multiobjetivo* de f em x se

$$\tilde{v} := \operatorname{argmin}_{\substack{v \in \mathbb{R}^n \\ x+v \in C}} \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v + \frac{1}{2}\|v\|^2. \quad (3.2)$$

Na seção de análise da direção de Cauchy multiobjetivo será mostrado que (3.2) sempre tem uma única solução.

3.1 Análise da direção de Cauchy multiobjetivo

Agora o objetivo será demonstrar que a direção de Cauchy multiobjetivo provém de uma aplicação regular de direções, da Definição 5, e com isso obter os resultados do Teorema 1. Mas antes serão necessárias algumas definições e proposições.

A análise feita aqui é um pouco mais geral à apresentada em (GRANA DRUMMOND; SVAITER, 2015; FUKUDA; DRUMMOND, 2014), pois ela serve de base para outros métodos de direções factíveis de descida que serão analisados.

Definição 11. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C . Defina $\Lambda_f: C \rightarrow \mathbb{R}^n$, $\theta_f: C \rightarrow \mathbb{R}$ e $\phi_x: \mathbb{R}^n \rightarrow \mathbb{R}$, com $x \in C$, onde:

- $\phi_x(v) :=$ função contínua, sublinear (ver Definição 17 no Apêndice A) e que satisfaz:
 - i. $J_f(x)v \prec 0 \implies M \leq \phi_x(v) < 0$.
 - ii. $J_f(x)v \not\prec 0 \implies 0 \leq \phi_x(v) \leq M$,
onde $M := \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$.
 - iii. J_f contínua em $C \implies \forall \{x_k\}_{k \in \mathbb{N}} \subseteq C, x_k \rightarrow x$:

$$\forall v \in \mathbb{R}^n: \quad \phi_{x_k}(v) \rightarrow \phi_x(v)$$

(converge pontualmente)

- $\Lambda_f(x) := \operatorname{argmin}_{\substack{v \in \mathbb{R}^n \\ x+v \in C}} \phi_x(v) + \frac{1}{2}\|v\|^2$.
- $\theta_f(x) := \phi_x(\Lambda_f(x)) + \frac{1}{2}\|\Lambda_f(x)\|^2$.

Na Definição 11, se $\phi_x(v) = \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$ então Λ_f define a direção de Cauchy multiobjetivo.

Proposição 9. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C e $x \in C$.

Então, $\Lambda_f(x)$ está bem definido.

Demonstração. Seja $h: (C - x) \rightarrow \mathbb{R}$ uma função definida como

$$h(v) := \phi_x(v) + \frac{1}{2}\|v\|^2,$$

onde $(C - x) := \{v \in \mathbb{R}^n : x + v \in C\}$.

Note que $(C - x)$ é a translação de um conjunto convexo fechado, e portanto também é convexo fechado.

Afirmção: h é coerciva em $(C - x)$. Se $(C - x)$ é limitado, h é coerciva. Se $(C - x)$ é ilimitado,

$$\begin{aligned} \lim_{\substack{\|v\| \rightarrow +\infty \\ v \in (C-x)}} h(v) &= \lim_{\substack{\|v\| \rightarrow +\infty \\ v \in (C-x)}} \phi_x(v) + \frac{1}{2}\|v\|^2 \\ &\geq \lim_{\substack{\|v\| \rightarrow +\infty \\ v \in (C-x)}} \left(- \max_{j \in \{1, \dots, m\}} \|v\| \|\nabla f_j(x)\| + \frac{1}{2}\|v\|^2 \right) \rightarrow +\infty. \end{aligned}$$

Afirmção: h é estritamente convexa. Sejam $u, v \in (C - x)$, $u \neq v$, e $\lambda \in (0, 1)$. Por ϕ_x ser sublinear e $(C - x)$ convexo,

$$\begin{aligned} \phi_x(\lambda v + (1 - \lambda)u) &\leq \phi_x(\lambda v) + \phi_x((1 - \lambda)u) \\ &= \lambda \phi_x(v) + (1 - \lambda) \phi_x(u). \end{aligned} \quad (3.3)$$

Além disso,

$$\begin{aligned} \|\lambda v + (1 - \lambda)u\|^2 &= \|u + \lambda(v - u)\|^2 \\ &= \|u\|^2 + \lambda^2\|v - u\|^2 + 2\lambda u^T(v - u) \\ &< \|u\|^2 + \lambda\|v - u\|^2 + 2\lambda u^T(v - u) \\ &= \|u\|^2 + \lambda\|v\|^2 + \lambda\|u\|^2 + 2\lambda u^T(-u) \\ &= \|u\|^2 + \lambda(\|v\|^2 - \|u\|^2) \\ &= \lambda\|v\|^2 + (1 - \lambda)\|u\|^2, \end{aligned}$$

portanto,

$$\|\lambda v + (1 - \lambda)u\|^2 < \lambda\|v\|^2 + (1 - \lambda)\|u\|^2. \quad (3.4)$$

Usando (3.3) e (3.4), conclui-se que, $\forall u, v \in C, u \neq v$:

$$\begin{aligned} h(\lambda v + (1 - \lambda)u) &= \phi_x(\lambda v + (1 - \lambda)u) + \|\lambda v + (1 - \lambda)u\|^2 \\ &< \lambda\phi_x(v) + (1 - \lambda)\phi_x(u) + \lambda\|v\|^2 + (1 - \lambda)\|u\|^2 \\ &= \lambda h(v) + (1 - \lambda)h(u), \quad \forall \lambda \in (0, 1). \end{aligned}$$

Como h é contínua e coerciva em $(C - x)$ convexo, pela Proposição 6, h admite um minimizador global em $(C - x)$. E como h é estritamente convexa, pela Proposição 7, o minimizador é único.

Então $\Lambda_f(x)$, que é o minimizador de h , está bem definido. ■

Proposição 10. *Sejam $C \subseteq \mathbb{R}^n$ convexo e fechado, $x \in C$ e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C . Então*

- i. $\phi_x(\Lambda_f(x)) \leq -\|\Lambda_f(x)\|^2$
- ii. $-\left(\max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|\right)^2 \leq 2\theta_f(x) \leq -\|\Lambda_f(x)\|^2$
- iii. $\|\Lambda_f(x)\| \leq \|J_f(x)\|$.

Demonstração. Essa demonstração será dividida em dois casos.

Caso $\|\Lambda_f(x)\| = 0$. Então todas afirmações são triviais:

$$\text{i. } \phi_x(0) = 0 \leq -\|\Lambda_f(x)\|^2 = 0.$$

- ii. $-\left(\max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|\right)^2 \leq 0 = \phi_x(0) + 0^2 = 2\theta_f(x).$
 $2\theta_f(x) = -\|\Lambda_f(x)\|^2.$
- iii. $\|\Lambda_f(x)\| = 0 \leq \|J_f(x)\|.$

Caso $\|\Lambda_f(x)\| \neq 0$. Então:

- i. Seja $\xi: [0, 1] \rightarrow \mathbb{R}$ definida como

$$\begin{aligned}\xi(t) &:= \phi_x(t\Lambda_f(x)) + \frac{1}{2}\|t\Lambda_f(x)\|^2 \\ &= t\phi_x(\Lambda_f(x)) + \frac{t^2}{2}\|\Lambda_f(x)\|^2.\end{aligned}$$

Pela definição, $\Lambda_f(x)$ é o minimizador de $\phi_x(v) + \frac{1}{2}\|v\|^2$ para $v \in \mathbb{R}^n$ tal que $x + v \in C$. Como C é convexo então $x + tv \in C, \forall t \in [0, 1]$, logo, $\xi(1)$ é o mínimo de ξ . Assim, se $t \in (0, 1)$,

$$\begin{aligned}\xi(t) - \xi(1) &\geq 0 \\ \implies (t-1)\phi_x(\Lambda_f(x)) + \frac{t^2-1}{2}\|\Lambda_f(x)\|^2 &\geq 0 \\ \implies \phi_x(\Lambda_f(x)) + \frac{t+1}{2}\|\Lambda_f(x)\|^2 &\leq 0 \\ \implies \lim_{t \rightarrow 1^-} \phi_x(\Lambda_f(x)) + \frac{t+1}{2}\|\Lambda_f(x)\|^2 &\leq 0 \\ \implies \phi_x(\Lambda_f(x)) &\leq -\|\Lambda_f(x)\|^2.\end{aligned}$$

- ii. Com o item i.:

$$\begin{aligned}2\theta_f(x) &= 2\phi_x(\Lambda_f(x)) + \|\Lambda_f(x)\|^2 \\ &\leq -2\|\Lambda_f(x)\|^2 + \|\Lambda_f(x)\|^2 \\ &\leq -\|\Lambda_f(x)\|^2.\end{aligned}$$

Para a outra desigualdade, note que

$$\begin{aligned}
2\theta_f(x) &= 2 \min_{\substack{v \in \mathbb{R}^n \\ x+v \in C}} \phi_x(v) + \frac{1}{2}\|v\|^2. \\
&\geq 2 \min_{\substack{v \in \mathbb{R}^n \\ x+v \in C}} \left(- \max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\| \|v\| + \frac{1}{2}\|v\|^2 \right).
\end{aligned}$$

Como o problema de minimização obtido depende somente da norma de v , então vale que:

$$2\theta_f(x) \geq 2 \min_{t \in \mathbb{R}} \left(-t \max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\| + \frac{t^2}{2} \right). \quad (3.5)$$

Como a função de t é uma função de segundo grau com concavidade voltada para cima, o minimizador é obtido igualando a derivada da função a zero:

$$\begin{aligned}
0 &= - \max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\| + t \\
\implies t &= \max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|.
\end{aligned}$$

Utilizando o t que minimiza (3.5) chega-se em:

$$\begin{aligned}
2\theta_f(x) &\geq -2 \left(\max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\| \right)^2 \\
&\quad + \left(\max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\| \right)^2 \\
&= - \left(\max_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\| \right)^2.
\end{aligned}$$

iii. Pela Definição 2 e pelos itens i. e ii.,

$$\|J_f(x)\| = \max_{v \in \mathbb{R}^n \setminus \{0\}} \frac{\|J_f(x)v\|}{\|v\|}$$

$$\begin{aligned}
&\geq \frac{1}{\|\Lambda_f(x)\|} \|J_f(x)\Lambda_f(x)\| \\
&= \frac{1}{\|\Lambda_f(x)\|} \sqrt{\sum_{j=1}^m (\nabla f_j(x)^T \Lambda_f(x))^2} \\
&\geq \frac{1}{\|\Lambda_f(x)\|} \max_{j \in \{1, \dots, m\}} |\nabla f_j(x)^T \Lambda_f(x)|.
\end{aligned}$$

Como

$$\begin{aligned}
&\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T \Lambda_f(x) \leq \phi_x(\Lambda_f(x)) < 0 \\
\implies &\max_{j \in \{1, \dots, m\}} |\nabla f_j(x)^T \Lambda_f(x)| \geq |\phi_x(\Lambda_f(x))| > 0,
\end{aligned}$$

conclui-se que,

$$\|J_f(x)\| \geq \frac{|\phi_x(\Lambda_f(x))|}{\|\Lambda_f(x)\|} \geq \frac{\|\Lambda_f(x)\|^2}{\|\Lambda_f(x)\|} = \|\Lambda_f(x)\|.$$

■

O resultado seguinte mostra outras caracterizações para pontos não críticos e direções factíveis de descida.

Proposição 11. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C e $x \in C$. Então são equivalentes:*

- i. x não é ponto crítico.*
- ii. $\theta_f(x) < 0$*
- iii. $\Lambda_f(x) \neq 0$*
- iv. $\Lambda_f(x)$ é uma direção factível de descida em x .*

Demonstração. (i. \implies ii.) Por hipótese, $\exists \tilde{v} \in \mathbb{R}^n$ direção factível de descida em x e, portanto, $\phi_x(\tilde{v}) < 0$.

Defina $h: \mathbb{R} \rightarrow \mathbb{R}$ como

$$h(t) := \phi_x(t\tilde{v}) + \frac{1}{2}\|t\tilde{v}\|^2 = t\phi_x(\tilde{v}) + \frac{t^2}{2}\|\tilde{v}\|^2.$$

Note que h é uma função de segundo grau com concavidade para cima e com raízes $t = 0$ e $t = -2\frac{\phi_x(\tilde{v})}{\|\tilde{v}\|^2} > 0$, pois $\phi_x(\tilde{v}) < 0$. Portanto, $\exists \tilde{t} \in (0, 1)$ tal que $h(\tilde{t}) < 0$.

Além disso, $x + \tilde{t}\tilde{v} \in C$, logo, $\tilde{t}\tilde{v}$ é direção factível de descida.

Como por construção $\theta_f(x) < h(\tilde{t})$ então $\theta_f(x) < 0$.

(ii. \implies iii.) Pela Proposição 10, $\theta_f(x) \leq 0$.

Além disso, $\|\Lambda_f(x)\| = 0 \implies \theta_f(x) = \phi_x(0) + \|0\|^2 = 0$.

Portanto, $\theta_f(x) \neq 0 \implies \theta_f(x) < 0 \implies \|\Lambda_f(x)\| \neq 0$.

(iii. \implies iv.) Da Proposição 10, e da definição de ϕ_x ,

$$\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T \Lambda_f(x) \leq \phi_x(\Lambda_f(x)) \leq -\|\Lambda_f(x)\|^2 < 0,$$

e, portanto, $\Lambda_f(x)$ é direção factível de descida.

(iv. \implies i.) Por definição, como existe direção factível de descida em x então x não é ponto crítico. ■

Proposição 12. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C .*

Então, Λ_f é limitada sob compactos, isto é,

$$\sup_{x \in K} \|\Lambda_f(x)\| < +\infty, \quad \forall K \subseteq C \text{ compacto.}$$

Demonstração. Pela Proposição 10, $\forall x \in C$, $\|\Lambda_f(x)\| \leq \|J_f(x)\|$.

Por hipótese, J_f é contínua em C .

Portanto, restrito a um compacto, $J_f(K)$ é limitado. Consequentemente, $\Lambda_f(K)$ é limitado. ■

Proposição 13. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C .*

Então θ_f é semicontínua superiormente, isto é, toda sequência $\{x_k\}_{k \in \mathbb{N}} \subseteq C$ tal que $x_k \rightarrow x \in C$ satisfaz

$$\limsup_{k \rightarrow +\infty} \theta_f(x_k) \leq \theta_f(x).$$

Demonstração. Sejam $x \in C$ e $\{x_k\}_{k \in \mathbb{N}} \subseteq C$ tal que $x_k \rightarrow x$.

Note que

$$\Lambda_f(x) + x \in C \implies (\Lambda_f(x) + x - x_k) + x_k \in C, \quad \forall k \in \mathbb{N}. \quad (3.6)$$

Então,

$$\begin{aligned} \theta_f(x_k) &= \phi_{x_k}(\Lambda_f(x_k)) + \frac{1}{2} \|\Lambda_f(x_k)\|^2 \\ &= \phi_{x_k}(\Lambda_f(x) + x - x_k) + \frac{1}{2} \|\Lambda_f(x) + x - x_k\|^2 \\ &\leq \phi_{x_k}(\Lambda_f(x)) + \phi_{x_k}(x - x_k) \\ &\quad + \frac{1}{2} \|\Lambda_f(x)\|^2 + \frac{1}{2} \|x - x_k\|^2 + \Lambda_f(x)^T (x - x_k). \end{aligned} \quad (3.7)$$

Como J_f é contínua em C então $\phi_{x_k} \rightarrow \phi_x$ pontualmente:

$$\lim_{k \rightarrow +\infty} \phi_{x_k}(\Lambda_f(x)) \rightarrow \phi_x(\Lambda_f(x)).$$

Ainda por J_f ser contínua e a sequência contida em um compacto então J_f é limitada na sequência, portanto,

$$\begin{aligned}\phi_{x_k}(x_k - x) &\leq \max_{j \in \{1, \dots, m\}} \|\nabla f_j(x_k)\| \|x_k - x\| \implies \\ \limsup_{k \rightarrow +\infty} \phi_{x_k}(x_k - x) &\leq \limsup_{k \rightarrow +\infty} \max_{j \in \{1, \dots, m\}} \|\nabla f_j(x_k)\| \|x_k - x\| = 0.\end{aligned}$$

Por isso, tomando o limite dos dois lados da inequação (3.7):

$$\limsup_{k \rightarrow +\infty} \theta_f(x_k) \leq \phi_x(\Lambda_f(x)) + \frac{1}{2} \|\Lambda_f(x)\|^2 = \theta_f(x).$$

■

Proposição 14. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C .*

Então Λ_f é aplicação regular de direções da Definição 5.

Demonstração. Pela Proposição 11, x não é crítico se e somente se $\Lambda_f(x)$ é direção factível de descida e $\theta_f(x) < 0$. Pela mesma proposição, x é ponto crítico se e somente se $\Lambda_f(x)$ é nulo, que junto com a Proposição 10, implica que $\theta_f(x)$ também é nulo.

Portanto, a condição (i.) de aplicação regular de direções está satisfeita.

Como J_f é contínua, a Proposição 12 garante que Λ_f é limitado sob compactos.

Portanto, a condição (ii.) de aplicação regular de direções está satisfeita.

Suponha que exista uma sequência $\{x_k\}_{k \in \mathbb{N}} \subseteq C$ de pontos não críticos convergindo para um ponto $x \in C$ não crítico tal que

$$\limsup_{k \rightarrow +\infty} J_f(x_k) \Lambda(x_k) \neq 0.$$

Como cada x_k é não crítico,

$$\max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T \Lambda(x_k) \leq \phi_{x_k}(\Lambda(x_k)) \leq \theta_f(x_k), \quad \forall k \in \mathbb{N},$$

e pela Proposição 13, θ_f é semicontínua superiormente. Logo,

$$\begin{aligned} & \limsup_{k \rightarrow +\infty} J_f(x_k) \Lambda(x_k) \not\leq 0 \\ \implies & 0 \leq \limsup_{k \rightarrow +\infty} \max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T \Lambda(x_k) \\ & \leq \limsup_{k \rightarrow +\infty} \theta_f(x_k) \leq \theta_f(x), \end{aligned}$$

o que implica que x é ponto crítico, contrariando a hipótese.

Então, a condição (iii.) de aplicação regular de direções está satisfeita. ■

Agora as funções Λ_f , ϕ_x e θ_f serão relacionadas com a direção de Cauchy multiobjetivo.

Lema 4. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C . Então para*

$$\phi_x(v) = \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v, \quad \forall x \in C,$$

todas condições da Definição 11 são satisfeitas.

E neste caso, Λ_f é a direção de Cauchy multiobjetivo.

Demonstração. A função ϕ_x é contínua, por ser o máximo de funções lineares.

Sejam $u, v \in \mathbb{R}^n$ e $\lambda > 0$. Então ϕ_x é subaditiva pois:

$$\phi_x(u + v) = \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T (u + v)$$

$$\begin{aligned}
&\leq \left(\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T u \right) + \left(\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v \right) \\
&= \phi_x(u) + \phi_x(v),
\end{aligned}$$

e homogênea positiva de grau 1:

$$\begin{aligned}
\phi_x(\lambda v) &= \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T \lambda u \\
&= \lambda \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v = \lambda \phi_x(v).
\end{aligned}$$

Portanto, ϕ_x é sublinear e contínua.

Agora será mostrado as três condições que ϕ_x deve satisfazer, para $M := \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$:

- i. $J_f(x)v \prec 0 \implies M = \phi_x(v) < 0$.
- ii. $J_f(x)v \not\prec 0 \implies 0 \leq \phi_x(v) = M$.
- iii. Suponha J_f contínua em C e seja $\{x_k\}_{k \in \mathbb{N}} \subseteq C$ tal que $x_k \rightarrow x$. Para mostrar a convergência pontual, fixe $v \in \mathbb{R}^n$. Como J_f é contínua então o máximo dos produtos internos dos gradientes por um v constante também é contínua. Portanto, $\phi_{x_k}(v) \rightarrow \phi_x(v)$.

Dessa forma,

$$\Lambda_f(x) = \operatorname{argmin}_{\substack{v \in \mathbb{R}^n \\ x+v \in C}} \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v + \frac{1}{2} \|v\|^2,$$

que é a direção de Cauchy multiobjetivo. ■

Usando o Lema 4, conclui-se que a direção de Cauchy multiobjetivo está bem definida.

Além disso, pela Proposição 14, a direção de Cauchy multiobjetivo é uma aplicação regular de direções, então as condições do Teorema 1 são satisfeitas.

A direção de Cauchy multiobjetivo também é aplicação fortemente regular de direções. Para demonstrar isso será necessário um teorema e uma proposição.

Teorema 3. *Sejam $V \subseteq \mathbb{R}^n$ convexo, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ convexa e diferenciável, e $h: \mathbb{R}^n \rightarrow \mathbb{R}$ definida como*

$$h(v) = \max_{j \in \{1, \dots, m\}} g_j(v).$$

Suponha que $\tilde{v} \in V$ seja minimizador global de h em V . Então

$$\exists w \in [0, 1]^m : \begin{cases} \mathbf{1}^T w = 1 \\ w^T g(\tilde{v}) = h(\tilde{v}) \\ w^T J_g(\tilde{v})(v - \tilde{v}) \geq 0, \quad \forall v \in V. \end{cases}$$

Demonstração. Pelo Teorema de Danskin (BERTSEKAS, 2003), se uma função $h: \mathbb{R}^n \rightarrow \mathbb{R}$ é o máximo de m funções $g_j: \mathbb{R}^n \rightarrow \mathbb{R}$, $\forall j \in \{1, \dots, m\}$, diferenciáveis e convexas então o *subdiferencial* de h em $v \in \mathbb{R}^n$, denotado $\partial h(v)$, é dado por

$$\partial h(v) = \{J_g(v)^T w : w \in [0, 1]^m \wedge \mathbf{1}^T w = 1 \wedge w^T g(v) = h(v)\}.$$

Por uma proposição na mesma referência, dado uma função $h: \mathbb{R}^n \rightarrow \mathbb{R}$ convexa, um subconjunto convexo $V \subseteq \mathbb{R}^n$ e $\tilde{v} \in V$ minimizador global de h em V então

$$\exists d \in \partial h(\tilde{v}) : \quad d^T(v - \tilde{v}) \geq 0, \quad \forall v \in V,$$

que conclui o teorema.

Mais detalhes em (BERTSEKAS, 2003). ■

Proposição 15. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C , $x \in C$ e $\tilde{v} \in \mathbb{R}^n$ direção de Cauchy multiobjetivo em x . Então,*

$$\exists w \in [0, 1]^m: \mathbf{1}^T w = 1 \quad \wedge \quad \tilde{v} = \Pi_C(x - J_f(x)^T w) - x.$$

Demonstração. Defina a função $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ como

$$g(v) := J_f(x)v + \frac{1}{2}\|v\|^2 \mathbf{1}^m.$$

A função g está nas condições do Teorema 3, pois é diferenciável e, sendo soma de funções convexas, também é uma função convexa (no sentido da Definição 19). Defina

$$h(v) = \max_{j \in \{1, \dots, m\}} g_j(v).$$

Defina $(C - x) := \{v \in \mathbb{R}^n: x + v \in C\}$. Esse conjunto é convexo e fechado por C ser convexo e fechado.

Note que o minimizador de h em $(C - x)$ é a direção de Cauchy multiobjetivo, ou seja, \tilde{v} .

Pelo Teorema 3, se \tilde{v} é o minimizador de h em $(C - x)$ então $\exists w \in [0, 1]^m$ onde $\mathbf{1}^T w = 1$ e

$$\begin{cases} w^T g(\tilde{v}) = h(\tilde{v}) \\ w^T J_g(\tilde{v})(v - \tilde{v}) \geq 0, \quad \forall v \in (C - x). \end{cases}$$

As duas condições precisam ser satisfeitas. Será usada apenas a segunda, para definir uma condição necessária para \tilde{v} . Assim, $\forall v \in (C - x)$,

$$\begin{aligned} w^T J_g(\tilde{v})(v - \tilde{v}) &\geq 0 \\ \iff (J_f(x)^T w + \tilde{v})^T (v - \tilde{v}) &\geq 0 \end{aligned}$$

$$\Longleftrightarrow (-J_f(x)^T w - \tilde{v})^T (v - \tilde{v}) \leq 0.$$

Pela Proposição 8, como a inequação acima se aplica $\forall v \in (C - x)$ então

$$\tilde{v} = \Pi_{(C-x)}(-J_f(x)^T w) = \Pi_C(x - J_f(x)^T w) - x.$$

■

Proposição 16. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C .*

Então a direção de Cauchy multiobjetivo define uma aplicação fortemente regular de direções, da Definição 9.

Demonstração. Seja $x \in C$.

Pelo Lema 4, para um ϕ_x específico (definido no Lema 4), e pela Proposição 14, a direção de Cauchy multiobjetivo define uma aplicação regular de direções.

Suponha f convexa.

Pela Proposição 15, se $\tilde{v} \in \mathbb{R}^n$ é direção de Cauchy multiobjetivo em $x \in C$ então $\exists w \in [0, 1]^m$ tal que

$$\mathbf{1}^T w = 1 \quad \wedge \quad \tilde{v} = \Pi_C(x - J_f(x)^T w) - x.$$

Logo, direção de Cauchy multiobjetivo define uma aplicação fortemente regular de direções. ■

Pela Proposição 16, o Algoritmo 2 com a direção de Cauchy multiobjetivo satisfaz as condições dos Teoremas 1 e 2.

Um problema equivalente a (3.2) que também resulta na direção de Cauchy multiobjetivo é descrito abaixo, na Proposição 17. Do ponto de vista prático, o problema (3.8) desta proposição

é mais conveniente para implementação, enquanto o problema (3.2), da definição de direção de Cauchy multiobjetivo, é mais conveniente para análise.

Proposição 17. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C e $x \in C$.*

Sejam $\tilde{v} \in \mathbb{R}^n$ e $\tilde{\tau} \in \mathbb{R}$ tais que

$$(\tilde{v}, \tilde{\tau}) = \begin{cases} \underset{\substack{(v, \tau) \in \mathbb{R}^n \times \mathbb{R} \\ x+v \in C}}{\operatorname{argmin}} & \tau + \frac{1}{2}\|v\|^2 \\ \text{sujeito a} & J_f(x)v - \tau \mathbf{1}^m \preceq 0. \end{cases} \quad (3.8)$$

Então \tilde{v} é a direção de Cauchy multiobjetivo em x .

Demonstração. Note a equivalência

$$J_f(x)v - \tau \mathbf{1}^m \preceq 0 \iff \tau \geq \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v.$$

Como τ é minimizado, $\tau = \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$. Assim,

$$\tilde{v} = \underset{\substack{v \in \mathbb{R}^n \\ x+v \in C}}{\operatorname{argmin}} \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v + \frac{1}{2}\|v\|^2.$$

Logo, \tilde{v} é direção de Cauchy multiobjetivo em x . ■

O problema (3.8) na Proposição 17 é um problema de programação quadrática, quando C pode ser definido por restrições lineares, cujos métodos de resolução são apresentados em diversas fontes, como (BERTSEKAS, 1999) e (NOCEDAL; WRIGHT, 1999).

3.2 Exemplos de uso

O método de Cauchy multiobjetivo é usado, nesta seção, para resolver os exemplos já vistos no Capítulo 1. Para o critério de parada, foi utilizado $J_f(x_k)v_k \succeq -10^{-8}$.

As Figuras 6, 7, 8 e 9 mostram as iterações do método de Cauchy multiobjetivo, respectivamente, nos problemas das Figuras 1, 2, 3 e 4.

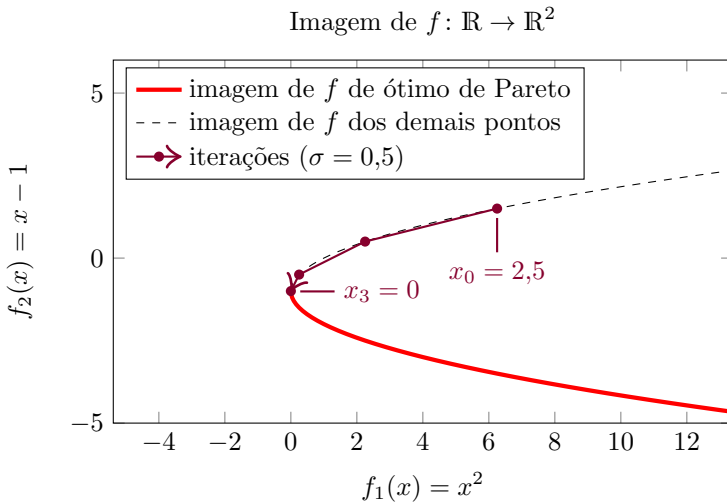


Figura 6 – A imagem de f e as iterações do método de Cauchy multiobjetivo. Com ponto inicial $x_0 = 2,5$, em 3 iterações atingiu-se $x_3 = 0$.

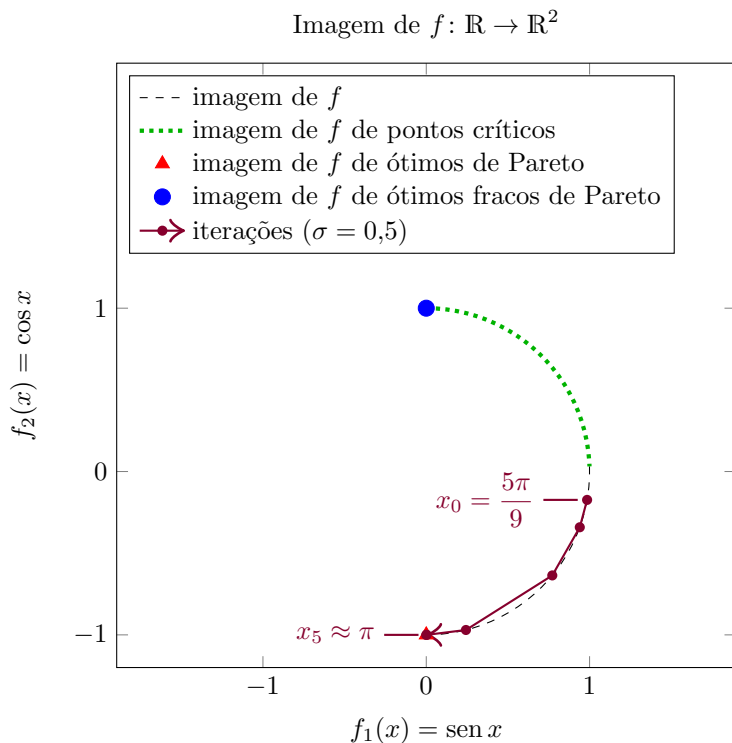


Figura 7 – A imagem de f e as iterações do método de Cauchy multiobjetivo. Com ponto inicial $x_0 = \frac{5}{9}\pi$, em 5 iterações atingiu-se $x_5 = 3,1416 \approx \pi$. Se o ponto inicial fosse ponto crítico não ótimo, isto é, se $x_0 \in \left[0, \frac{\pi}{2}\right]$, o algoritmo terminaria logo na primeira iteração.

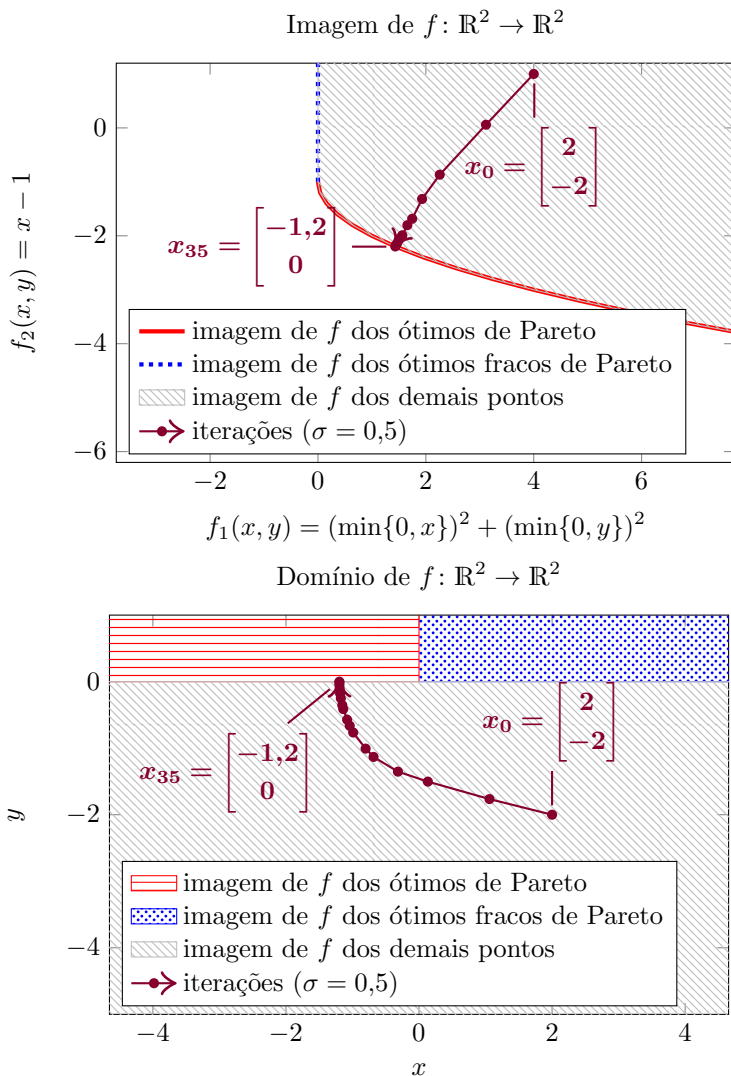


Figura 8 – Iterações do método de Cauchy multiobjetivo em f . Com ponto inicial x_0 , em 35 iterações atingiu-se x_{35} .

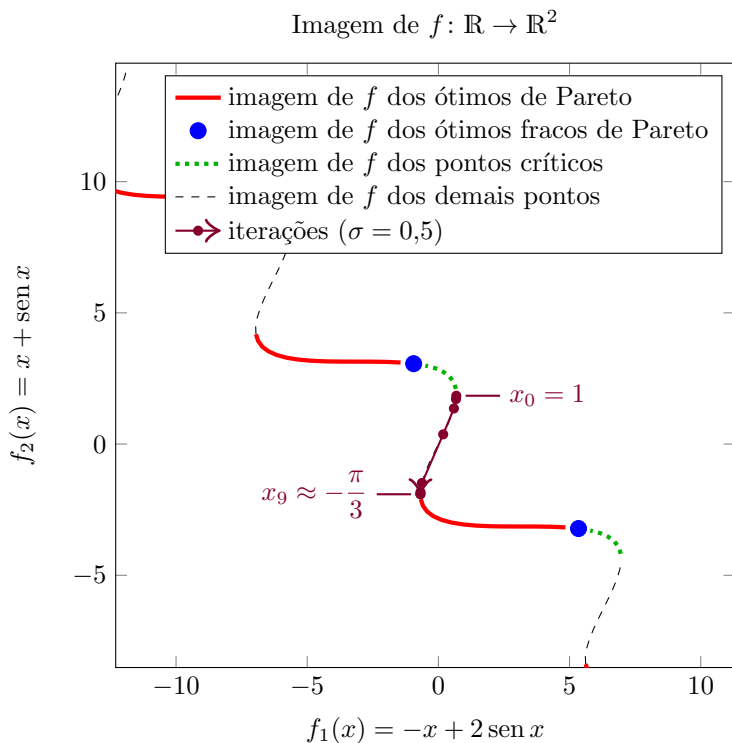


Figura 9 – As iterações do método de Cauchy multiobjetivo na imagem de f . Com ponto inicial $x_0 = 1$, em 9 iterações atingiu-se $x_9 = -1,047 \approx -\frac{\pi}{3}$. Se x_0 fosse um pouco maior, como $x_0 = \frac{\pi}{3} \approx 1,047$, ele já seria um ponto crítico não ótimo e o algoritmo terminaria logo na primeira iteração.

4 Outras direções de descida

4.1 Direção Central

A *direção central* é baseada na direção de Cauchy multiobjetivo, da Definição 10, mas com os gradientes normalizados, e o resultado é uma direção factível de descida próxima ao centro do cone formado pelos gradientes negativos.

A motivação para esta direção factível de descida é que a direção de Cauchy multiobjetivo dá prioridade para ficar na direção do gradiente negativo de menor norma. Se um gradiente tiver uma norma muito menor que os demais, a direção de Cauchy multiobjetivo será quase igual a este gradiente negativo. Mas todos gradientes recebem a mesma prioridade se estiverem normalizados.

Definição 12. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C .

Defina, $\forall x \in C$, $L_x \in \mathbb{R}$ como

$$L_x := \min_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|.$$

Seja $x \in C$.

Se $L_x = 0$, a *direção central* de f em x é $\tilde{v} = 0 \in \mathbb{R}^n$.

Se $L_x \neq 0$, a *direção central* de f em x é $\tilde{v} \in \mathbb{R}^n$ onde

$$\tilde{v} = \underset{\substack{v \in \mathbb{R}^n \\ x+v \in C}}{\operatorname{argmin}} \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T v}{\|\nabla f_j(x)\|} + \frac{1}{2} \|v\|^2. \quad (4.1)$$

Será mostrado que o ϕ_x , da Definição 11, pode ser escolhido de forma que Λ_f seja uma função da direção central.

Lema 5. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C . Defina, $\forall x \in C$, $\phi_x: \mathbb{R}^n \rightarrow \mathbb{R}$ como*

$$\phi_x(v) = \begin{cases} 0, & \text{se } L_x = 0, \\ \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T v}{\|\nabla f_j(x)\|}, & \text{se } L_x > 0, \end{cases}$$

onde L_x é da Definição 12.

Então todas condições da Definição 11 são satisfeitas usando o ϕ_x definido. E neste caso, Λ_f é a direção central.

Demonstração. A demonstração será dividida em dois casos.

Caso $L_x = 0$: então, ϕ_x é contínua e sublinear, pois

$$0 = \phi_x(u + v) \leq \phi(u) + \phi(v) = 0, \quad \forall u, v \in C,$$

$$0 = \phi_x(\lambda u) = \lambda \phi_x(u) = 0, \quad \forall u \in C, \forall \lambda > 0.$$

Agora será mostrado que ϕ_x satisfaz as três condições necessárias, para $M := \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$:

- i. Satisfaz trivialmente, pois $J_f(x)v \not\prec 0$.
- ii. $J_f(x)v \not\prec 0 \implies 0 = \phi_x(v) \leq M$.
- iii. Suponha J_f contínua em C . Para mostrar a convergência pontual, seja $\{x_k\}_{k \in \mathbb{N}} \subseteq C$, $x_k \rightarrow x$ e fixe $v \in \mathbb{R}^n$.

Se $L_{x_k} = 0$, $\phi_{x_k}(v) = 0$.

Se $L_{x_k} > 0$,

$$\begin{aligned}
 |\phi_{x_k}(v)| &= \left| \max_{j \in \{1, \dots, m\}} \frac{L_{x_k} \nabla f_j(x_k)^T v}{\|\nabla f_j(x_k)\|} \right| \\
 &\leq \max_{j \in \{1, \dots, m\}} \left| \frac{L_{x_k} \nabla f_j(x_k)^T v}{\|\nabla f_j(x_k)\|} \right| \\
 &\leq \max_{j \in \{1, \dots, m\}} \frac{L_{x_k} \|\nabla f_j(x_k)\| \|v\|}{\|\nabla f_j(x_k)\|} \\
 &= L_{x_k} \|v\| \\
 &\implies \lim_{k \rightarrow +\infty} |\phi_{x_k}(v)| \leq L_x \|v\| = 0 \\
 &\implies \lim_{k \rightarrow +\infty} \phi_{x_k}(v) = 0 = \phi_x(v),
 \end{aligned}$$

portanto, $\phi_{x_k} \rightarrow \phi_x$ pontualmente, neste caso.

Caso $L_x > 0$: então, ϕ_x é contínua, pois é o máximo de funções contínuas.

Sejam $u, v \in \mathbb{R}^n$ e $\lambda > 0$. Então ϕ_x é subaditiva:

$$\begin{aligned}
 \phi_x(u + v) &= \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T (u + v)}{\|\nabla f_j(x)\|} \\
 &\leq \left(\max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T u}{\|\nabla f_j(x)\|} \right) \\
 &\quad + \left(\max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T v}{\|\nabla f_j(x)\|} \right) \\
 &= \phi_x(u) + \phi_x(v),
 \end{aligned}$$

e homogênea positiva de grau 1: $\forall \lambda > 1$,

$$\begin{aligned}
 \phi_x(\lambda v) &= \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T \lambda v}{\|\nabla f_j(x)\|} \\
 &= \lambda \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T v}{\|\nabla f_j(x)\|} = \lambda \phi_x(v).
 \end{aligned}$$

Note que pela definição de L_x , tem-se que

$$0 < \frac{L_x}{\|\nabla f_j(x)\|} \leq 1, \quad \forall j \in \{1, \dots, m\}. \quad (4.2)$$

Assim, dado $M := \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$, será mostrado que ϕ_x satisfaz as três condições necessárias:

i. Por (4.2), $\forall j \in \{1, \dots, m\}$,

$$J_f(x)v \prec 0 \implies M \leq \frac{L_x M}{\|\nabla f_j(x)\|} \leq \phi_x(v) < 0.$$

ii. Supondo $J_f(x)v \not\prec 0$, e ainda por (4.2),

$$0 \leq \phi_x(v) \leq \max_{j \in \{1, \dots, m\}} \frac{L_x M}{\|\nabla f_j(x)\|} \leq M.$$

iii. Suponha J_f contínua em C . Para mostrar a convergência pontual, seja $\{x_k\}_{k \in \mathbb{N}} \subseteq C$, $x_k \rightarrow x$ e fixe $v \in \mathbb{R}^n$.

Como $L_x > 0$ e pela continuidade dos gradientes, $\exists N \in \mathbb{N}$ tal que se $k > N$ então $L_{x_k} > 0$, pois $L_x = \min_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|$ é contínua em x , pela continuidade de J_f . Além disso,

$$\phi_x(\lambda v) = \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T \lambda v}{\|\nabla f_j(x)\|},$$

que é o máximo de funções contínuas em x , portanto, contínua em x . Assim, $\phi_{x_k} \rightarrow \phi_x$ pontualmente.

Dessa forma, todas condições de ϕ_x da Definição 11 são satisfeitas e $\Lambda_f(x)$ define a direção central.

■

Usando o Lema 5 acima conclui-se que a direção central está bem definida.

Além disso, pela Proposição 14, a direção central é uma aplicação regular de direções, então as condições do Teorema 1 são satisfeitas.

A próxima proposição mostra que a direção central é também aplicação fortemente regular de direções.

Proposição 18. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C .*

Então a direção central define uma aplicação fortemente regular de direções, da Definição 9.

Demonstração. Seja $x \in C$.

Pelo Lema 5, a direção central pode ser definida como $\Lambda_f(x)$, para um ϕ_x específico (definido no Lema 5), e pela Proposição 14, Λ_f define uma aplicação regular de direções.

Portanto a direção central define uma aplicação regular de direções.

Suponha f convexa. Seja $\tilde{v} \in \mathbb{R}^n$ a direção central em x .

A demonstração será dividida em dois casos.

Caso $L_x = 0$ Então x é ponto crítico, pois $J_f(x)v \neq 0, \forall v \in \mathbb{R}^n$. Pela Proposição 11, $\Lambda_f(x) = 0$ e portanto $\tilde{v} = 0$. Por isso, para $w = 0$, todas condições de aplicação fortemente regular de direções são satisfeitas.

Caso $L_x > 0$ Essa demonstração é muito parecida com a Proposição 15.

Defina $A_x \in \mathbb{R}^{m \times m}$ uma matriz diagonal onde

$$(A_x)_{jj} := \frac{L_x}{\|\nabla f_j(x)\|}, \quad \forall j \in \{1, \dots, m\}. \quad (4.3)$$

Defina a função $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ como

$$g(v) := A_x J_f(x)v + \frac{1}{2}\|v\|^2 \mathbf{1}^m.$$

A função g está nas condições do Teorema 3, pois é diferenciável e, sendo soma de funções convexas, também é uma função convexa. Defina

$$h(v) = \max_{j \in \{1, \dots, m\}} g_j(v).$$

Defina $(C - x) := \{v \in \mathbb{R}^n : x + v \in C\}$. Esse conjunto é convexo e fechado por C ser convexo e fechado.

Note que o minimizador de h em $(C - x)$ é a direção central, ou seja, \tilde{v} .

Pelo Teorema 3, se \tilde{v} é o minimizador de h em $(C - x)$ então $\exists w \in [0, 1]^m$ onde $\mathbf{1}^T w = 1$ e

$$\begin{cases} w^T g(\tilde{v}) = h(\tilde{v}) \\ w^T J_g(\tilde{v})(v - \tilde{v}) \geq 0, \quad \forall v \in (C - x). \end{cases}$$

Para a segunda condição ser satisfeita, então $\forall v \in (C - x)$,

$$\begin{aligned} & w^T J_g(\tilde{v})(v - \tilde{v}) \geq 0 \\ \iff & (J_f(x)^T A_x w + \tilde{v})^T (v - \tilde{v}) \geq 0 \\ \iff & (-J_f(x)^T A_x w - \tilde{v})^T (v - \tilde{v}) \leq 0. \end{aligned}$$

Pela Proposição 8, como a inequação acima se aplica $\forall v \in (C - x)$ então

$$\tilde{v} = \Pi_{(C-x)}(-J_f A_x(x)^T w) = \Pi_C(x - J_f(x)^T A_x w) - x.$$

Seja $\tilde{w} = A_x w$. Como $0 < (A_x)_{jj} \leq 1, \forall j \in \{1, \dots, m\}$ então $\mathbf{1}^T \tilde{w} \leq 1, \tilde{w} \succeq 0$ e

$$\begin{aligned} \Pi_C(x - J_f(x)^T A_x w) - x &= \Pi_C(x - J_f(x)^T \tilde{w}) - x \\ \implies \tilde{v} &= \Pi_C(x - J_f(x)^T \tilde{w}) - x. \end{aligned}$$

Então, a direção central define uma aplicação fortemente regular de direções. ■

Pela Proposição 18, o Algoritmo 2 com a direção central satisfaz as condições dos Teoremas 1 e 2.

Um problema equivalente a (4.1) que também resulta na direção central é descrito abaixo, que é uma versão mais conveniente para implementação.

Proposição 19. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C e $x \in C$. Defina $A_x \in \mathbb{R}^{m \times m}$ matriz diagonal como*

$$(A_x)_{jj} := \begin{cases} 0, & \text{se } L_x = 0, \\ \frac{L_x}{\|\nabla f_j(x)\|}, & \text{se } L_x \neq 0, \end{cases}$$

onde L_x é da Definição 12.

Sejam $\tilde{v} \in \mathbb{R}^n$ e $\tilde{\tau} \in \mathbb{R}$ tais que

$$(\tilde{v}, \tilde{\tau}) = \begin{cases} \underset{\substack{(v, \tau) \in \mathbb{R}^n \times \mathbb{R} \\ x+v \in C}}{\operatorname{argmin}} & \tau + \frac{1}{2} \|v\|^2 \\ \text{sujeito a} & A_x J_f(x) v - \tau \mathbf{1}^m \preceq 0. \end{cases} \quad (4.4)$$

Então \tilde{v} é a direção central em x .

Demonstração. Caso $L_x = 0$ então $\tilde{v} = 0 \in \mathbb{R}^n$ é solução de (4.4) e, portanto, direção central em x .

Suponha $L_x \neq 0$. Neste caso, note a equivalência:

$$A_x J_f(x)v - \tau \mathbf{1}^m \preceq 0 \iff \tau \geq \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T v}{\|\nabla f_j(x)^T v\|}.$$

Por τ ser minimizado, $\tau = \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T v}{\|\nabla f_j(x)^T v\|}$. Assim,

$$\tilde{v} = \underset{\substack{v \in \mathbb{R}^n \\ x+v \in C}}{\operatorname{argmin}} \max_{j \in \{1, \dots, m\}} \frac{L_x \nabla f_j(x)^T v}{\|\nabla f_j(x)^T v\|} + \frac{1}{2} \|v\|^2.$$

Logo, \tilde{v} é a direção central em x . ■

4.2 Direção Linear

A *direção linear* é baseada na direção de Cauchy multiobjetivo, da Definição 10, mas em vez de usar uma norma ao quadrado para limitar a função a ser minimizada, ela usa apenas uma restrição linear da norma infinito da direção (FLIEGE; SVAITER, 2000). Desta forma, o problema se torna minimizar uma função linear sujeita a restrições lineares, desde que o conjunto viável possa ser expresso apenas por restrições lineares. Neste caso, métodos de programação linear podem ser usados (NOCEDAL; WRIGHT, 1999).

Definição 13. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C .

Defina $\Delta_x \in \mathbb{R}$ satisfazendo

$$\begin{aligned} \Delta_x &> 0, \quad \forall x \in C, \\ 0 &< \inf_{x \in C} \Delta_x, \\ \sup_{x \in K} \Delta_x &< +\infty, \quad \forall K \subseteq C \text{ compacto.} \end{aligned}$$

Seja $x \in C$ e defina $u \in \mathbb{R}^n$ como

$$u = \begin{cases} \underset{\substack{v \in \mathbb{R}^n \\ x+v \in C}}{\operatorname{argmin}} & \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v \\ \text{sujeito a} & \|v\|_\infty \leq \Delta_x. \end{cases} \quad (4.5)$$

A *direção linear* de f em x é $\tilde{v} \in \mathbb{R}^n$ definida como

$$\tilde{v} = \begin{cases} u, & \text{se } J_f(x)u \prec 0, \\ 0 \in \mathbb{R}^n, & \text{se } J_f(x)u \not\prec 0. \end{cases} \quad (4.6)$$

O problema (4.5) está bem definido pois minimiza uma função contínua em um compacto não vazio. Porém, não existe unicidade na solução, como é o caso da direção de Cauchy multiobjetivo ou a direção central.

Vale notar que em (FLIEGE; SVAITER, 2000) a restrição usada é $\|v\|_\infty \leq 1$ em vez de $\|v\|_\infty \leq \Delta_x$. O uso de um Δ_x tem algumas vantagens na prática, discutidas no Capítulo 5.

Existe um conceito chamado *direção dominada* (HARADA et al., 2006). Uma direção factível de descida $v \in \mathbb{R}^n$ em $x \in C$ é dita *dominada* se existe $\tilde{v} \in \mathbb{R}^n$ direção factível de descida em x tal que $\|v\| = \|\tilde{v}\|$ e $J_f(x)\tilde{v} \prec J_f(x)v$, ou seja, teoricamente

todas funções componentes reduzem mais na direção \tilde{v} que na direção v .

Portanto, é natural preferir uma direção factível de descida *não dominada*, também chamada de *direção de Pareto*.

É possível mostrar que a direção de Cauchy multiobjetivo é não dominada no caso irrestrito ($C = \mathbb{R}^n$), assim como a direção central. Porém, a direção linear pode tomar direções dominadas, mesmo no caso irrestrito, e essa é uma desvantagem desta direção.

Mesmo assim, a direção linear provém de uma aplicação regular de direções, como será mostrado.

Proposição 20. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C e $\gamma: C \rightarrow \mathbb{R}$ onde $\gamma(x)$ é definida como uma direção linear em x , $\forall x \in C$.*

Então γ é aplicação regular de direções, da Definição 5.

Demonstração. Seja $x \in C$. Defina:

$$(C - x) := \{v \in \mathbb{R}^n: x + v \in C\}$$

$$V := \{v \in (C - x): \|v\|_\infty \leq \Delta_x\},$$

onde Δ_x é da Definição 13.

Portanto, V é compacto e convexo, pois é intersecção de compacto convexo com fechado convexo. Além disso, $0 \in V$.

Seja $h: V \rightarrow C$ definida como

$$h(v) = \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v.$$

Assim, o minimizador de h é u , de (4.5), pois o domínio de h é a restrição de (4.5).

Suponha que $v \in \mathbb{R}^n$ seja uma direção factível de descida em x . Então, existe $\lambda \in (0, 1)$ tal que $\|\lambda v\|_\infty \leq \Delta_x$. Como C

é convexo e $v \in (C - x)$ então $\lambda v \in (C - x)$, logo, $\lambda v \in V$ e $h(v) < 0$. E como u minimiza h então $h(u) \leq h(v) < 0$. Ou seja, se existir direção factível de descida, o minimizador de h será direção factível de descida. E por (4.6), $\gamma(x) = u$.

Quando não há direção factível de descida então $\forall v \in (C - x)$, $h(v) \geq 0$. Portanto, u , que minimiza h , não será direção factível de descida. E por (4.6), $\gamma(x) = 0 \in \mathbb{R}^n$.

A direção linear é limitada por Δ_x , que por sua vez é limitada sob compactos, então a direção linear é limitada sob compactos.

Logo, as condições (i.) e (ii.) de aplicação regular de direções estão satisfeitas.

Suponha que $\exists \{x_k\}_{k \in \mathbb{N}} \subseteq C$ de pontos não críticos convergindo para x não crítico.

Note que, $\forall v \in \mathbb{R}^n$,

$$\|v\| = \sqrt{\sum_{j=1}^m v_j^2} \geq \sqrt{\max_{j \in \{1, \dots, m\}} v_j^2} = \max_{j \in \{1, \dots, m\}} |v_j| = \|v\|_\infty$$

$$\implies \forall r \geq 0, \forall v \in \mathbb{R}^n: \|v\| \leq r \implies \|v\|_\infty \leq r.$$

Como $0 < \inf_{k \in \mathbb{N}} \Delta_{x_k}$ então $\exists \Delta > 0$ tal que, $\forall k \in \mathbb{N}$,

$$\Delta < \Delta_{x_k}$$

$$\implies \{v \in (C - x): \|v\| \leq \Delta\} \subseteq \{v \in (C - x): \|v\|_\infty \leq \Delta_{x_k}\}.$$

Portanto,

$$\begin{aligned} \gamma(x_k) &= \underset{\substack{v \in (C-x) \\ \|v\|_\infty \leq \Delta_{x_k}}}{\operatorname{argmin}} \max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T v \\ \implies \max_{j \in \{1, \dots, m\}} \nabla f_j(x_k) \gamma(x_k) &\leq \min_{\substack{v \in (C-x) \\ \|v\| \leq \Delta}} \max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T v. \end{aligned} \tag{4.7}$$

Seja Λ_f função da direção de Cauchy multiobjetivo, do Lema 4. Pela Proposição 12, Λ_f é limitado sob compactos, e como $\{x_k\}$ está contido em um compacto, seja $G > 0$ cota superior de $\{\|\Lambda_f(x_k)\| : k \in \mathbb{N}\}$. Pela Proposição 14, Λ_f é aplicação regular de direções, portanto

$$\limsup_{k \rightarrow +\infty} J_f(x_k) \Lambda_f(x_k) \prec 0 \implies \limsup_{k \rightarrow +\infty} J_f(x_k) \left(\frac{\Delta}{G} \Lambda_f(x_k) \right) \prec 0. \quad (4.8)$$

Seja $\{v_k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ tal que $v_k = \frac{\Delta}{G} \Lambda_f(x_k)$. Então,

$$\|v_k\| \leq \Delta \frac{\|\Lambda_f(x_k)\|}{G} \leq \Delta, \quad \forall k > N.$$

Então, por (4.7), $\forall k \in \mathbb{N}$,

$$\max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T \gamma(x_k) \leq \max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T v_k,$$

e juntando com (4.8),

$$\limsup_{j \rightarrow +\infty} J_f(x_k) \gamma(x_k) \preceq \limsup_{j \rightarrow +\infty} J_f(x_k) v_k \prec 0.$$

Logo, a condição (iii.) de aplicação regular de direções está satisfeita e a direção linear gera uma aplicação regular de direções. ■

Portanto, pela Proposição 20, o Algoritmo 2 com a direção linear satisfaz as condições do Teorema 1.

Um problema equivalente a (4.6) que também resulta na direção linear é descrito abaixo, que é uma versão mais conveniente para implementação.

Proposição 21. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, e $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C .*

Defina $\Delta_x \in \mathbb{R}$ satisfazendo

$$\begin{aligned} \Delta_x &> 0, \quad \forall x \in C, \\ 0 &< \inf_{x \in C} \Delta_x \\ \sup_{x \in K} \Delta_x &< +\infty, \quad \forall K \subseteq C \text{ compacto.} \end{aligned}$$

Seja $x \in C$ e defina $\tilde{v} \in \mathbb{R}^n$ e $\tilde{\tau} \in \mathbb{R}$ como

$$(u, \tilde{\tau}) = \begin{cases} \underset{\substack{(v, \tau) \in \mathbb{R}^n \times \mathbb{R} \\ x+v \in C}}{\operatorname{argmin}} \quad \tau \\ \text{sujeito a} \quad J_f(x)v - \tau \mathbf{1}^m \preceq 0, \\ \|v\|_\infty \leq \Delta_x. \end{cases} \quad (4.9)$$

Então $\tilde{v} \in \mathbb{R}^n$ definido como

$$\tilde{v} = \begin{cases} u, & \text{se } \tilde{\tau} < 0, \\ 0 \in \mathbb{R}^n, & \text{se } \tilde{\tau} \geq 0. \end{cases} \quad (4.10)$$

é uma direção linear em x .

Demonstração. Note a equivalência

$$J_f(x)v - \tau \mathbf{1}^m \preceq 0 \iff \tau \geq \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v.$$

Por τ ser minimizado, $\tau = \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$. Assim,

$$u = \begin{cases} \underset{\substack{v \in \mathbb{R}^n \\ x+v \in C}}{\operatorname{argmin}} \quad \max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v \\ \text{sujeito a} \quad \|v\|_\infty \leq \Delta_x. \end{cases}$$

Se $\tilde{\tau} \geq 0$ então $J_f(x)u \not\prec 0$.

Se $\tilde{\tau} < 0$ então $J_f(x)u \prec 0$.

Portanto, (4.10) é o mesmo que (4.6) da Definição 13, logo, \tilde{v} é uma direção linear de f em x . ■

O problema (4.9) na Proposição 21 é um problema de programação linear, quando C pode ser definido por restrições lineares.

4.3 Direção de Pareto descendente

Assim como a direção linear, a *direção de Pareto descendente* tem por objetivo encontrar uma direção factível de descida apenas resolvendo problema de programação linear (HARADA et al., 2006). A versão apresentada aqui é um pouco modificada, para facilitar a análise de convergência, mas a principal ideia da direção de Pareto descendente original é mantida.

A ideia principal é obter uma direção denominada *não dominada*, ou *direção de Pareto*, que é uma direção $\tilde{v} \in \mathbb{R}^n \setminus \{0\}$ que, $\forall v \in \mathbb{R}^n \setminus \{0\}$, $\frac{1}{\|v\|} J_f(x)v \preceq \frac{1}{\|\tilde{v}\|} J_f(x)\tilde{v} \implies \frac{1}{\|v\|} v = \frac{1}{\|\tilde{v}\|} \tilde{v}$. Para esta definição, em (HARADA et al., 2006), são considerados os gradientes também normalizados. Estas direções teoricamente são as que mais reduzem a função objetivo, pelo menos sob o ponto de vista da primeira derivada. Pode ser mostrado, segundo (HARADA et al., 2006), que uma direção é não dominada se e somente se for combinação convexa dos gradientes negativos normalizados.

Para encontrar essa combinação convexa, foi usado aqui a combinação cônica limitada: $w \in \mathbb{R}^m$ onde $w \succeq 0$ e $\mathbf{1}^T w \leq 1$, em que w são os pesos na combinação dos gradientes negativos

normalizados. Assim, a direção será da forma $v = -J_f(x)^T A_x w$, onde A_x é uma matriz diagonal positiva que faz com que todos gradientes de $J_f(x)$ tenham a mesma norma. Essa combinação cônica limitada é semelhante a combinação convexa, se os pesos forem normalizados para somar 1, e portanto a direção encontrada também será uma direção não dominada, a menos que os pesos somem zero. A vantagem de usar esta forma é aumentar o espaço de busca da direção factível de descida, e com isso pode-se demonstrar que a direção de Pareto descendente forma uma aplicação regular de direções.

Porém, devido à restrição do conjunto viável (isto é, $x+v \in C$), nem sempre existe direção factível de descida que é combinação cônica dos gradientes negativos (ver Figura 10). Quando não existir, a direção de Pareto descendente é tomada como sendo a direção linear.

Na versão original da direção de Pareto descendente, m problemas de programação linear eram resolvidos para determinar w , cada um minimizando uma componente w_j sujeito a que $-J_f(x)J_f(x)^T w \preceq 0$ e $w \in [0, 1]^m$, e isso resultava em m vetores w distintos. Então a direção final tomada usava uma combinação convexa entre os pesos encontrados. A versão modificada utilizada simplifica para haver somente um problema a ser resolvido.

Além disso, quando uma direção factível de descida não era encontrada na versão original, um problema semelhante à direção linear era resolvido, mas em vez de minimizar $\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v$, era minimizado $w^T v$ sujeito a $\max_{j \in \{1, \dots, m\}} \nabla f_j(x)^T v \preceq 0$, com w vetor aleatoriamente gerado, e diversas minimizações eram feitas com vetores w aleatórios diferentes para obter uma combinação cônica entre todas as direções v encontradas. Mais detalhes em (HARADA et al., 2006).

Exemplo de $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$

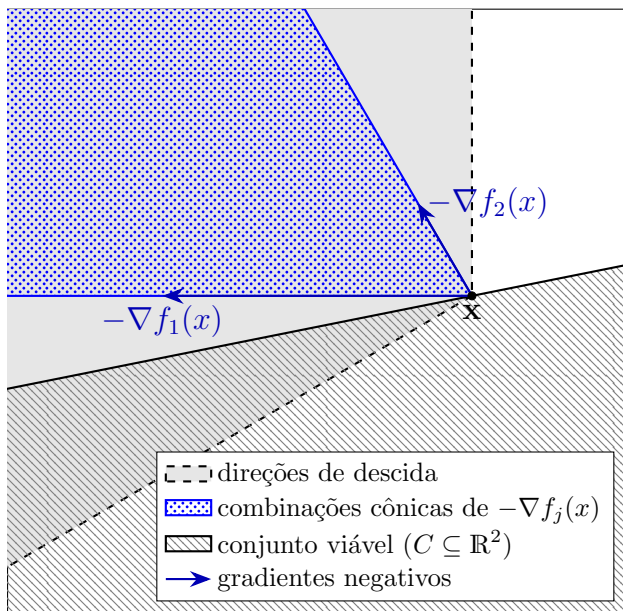


Figura 10 – Exemplo onde nenhuma direção de descida $v \in \mathbb{R}^2$ factível ($x + v \in C$) é combinação cônica dos gradientes negativos.

Definição 14. Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C . Fixe $\beta > 0$.

Defina, $\forall x \in C$, $L_x \in \mathbb{R}$ como

$$L_x := \min_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|.$$

Seja $x \in C$.

Se $L_x = 0$, a *direção de Pareto descendente* de f em x é definida como $\tilde{v} = 0 \in \mathbb{R}^n$.

Se $L_x \neq 0$, defina $A_x \in \mathbb{R}^{m \times m}$ matriz diagonal como

$$(A_x)_{jj} := \frac{L_x}{\|\nabla f_j(x)\|}, \quad j \in \{1, \dots, m\}.$$

Seja $\tilde{w} \in \mathbb{R}^n$ definida como

$$\tilde{w} = \begin{cases} \begin{array}{ll} \underset{\substack{w \in \mathbb{R}^m \\ x - J_f(x)^T A_x w \in C}}{\operatorname{argmin}} & \max_{j \in \{1, \dots, m\}} - \frac{L_x \nabla f_j(x)^T J_f(x)^T A_x w}{\|\nabla f_j(x)\|} \\ \text{sujeito a} & \mathbf{1}^T w \leq 1, \\ & w \succeq 0. \end{array} \end{cases} \quad (4.11)$$

A *direção de Pareto descendente* de f em x é $\tilde{v} \in \mathbb{R}^n$:

$$\tilde{v} = \begin{cases} -J_f(x)^T A_x \tilde{w}, & \text{se } -J_f(x) J_f(x)^T A_x \tilde{w} \preceq -\beta L_x^2 \mathbf{1}^m, \\ \text{direção linear}, & \text{se } -J_f(x) J_f(x)^T A_x \tilde{w} \not\preceq -\beta L_x^2 \mathbf{1}^m. \end{cases} \quad (4.12)$$

O problema (4.11) está bem definido pois consiste em minimizar uma função contínua em um compacto não vazio.

O β serve para evitar que a direção de Pareto descendente

obtenha direções factíveis de descida quase nulas próxima de pontos não críticos, que podem ocorrer dependendo do conjunto viável.

Proposição 22. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável em um aberto contendo C e $\gamma: C \rightarrow \mathbb{R}$ onde $\gamma(x)$ é definida como uma direção de Pareto descendente em x , $\forall x \in C$.*

Então γ é aplicação regular de direções, da Definição 5.

Demonstração. Será mostrado que todas condições da Definição 5 são satisfeitas.

Sejam $\tilde{w} \in \mathbb{R}^n$, $A_x \in \mathbb{R}^{m \times m}$, $L_x \in \mathbb{R}$ e $\beta > 0$ da Definição 14.

Condição (i.) Seja $x \in C$.

Se $L_x = 0$ então x é ponto crítico, e $\gamma(x) = 0 \in \mathbb{R}^n$.

Suponha $L_x \neq 0$. Neste caso, se $\gamma(x)$ é a direção linear então $\gamma(x) = 0 \in \mathbb{R}^n$ se x é ponto crítico e $\gamma(x)$ é uma direção factível de descida se x não é ponto crítico.

Suponha então $\gamma(x) = -J_f(x)^T A_x \tilde{w}$, de (4.12). Então $\gamma(x)$ é uma direção factível de descida, pois para $\gamma(x)$ ser definido desta forma em (4.12), é necessário que $J_f(x)\gamma(x) \preceq -\beta L_x^2 \mathbf{1}^m < 0$. Portanto, x não é ponto crítico.

Condição (ii.) Sejam $K \subseteq C$ um compacto e $x \in K$.

Se $\gamma(x)$ é direção linear, ele é limitado por algum número real, pois a direção linear é limitada sob compactos pela Proposição 20.

Se $\gamma(x)$ é da forma $\gamma(x) = -J_f(x)^T A_x \tilde{w}$, de (4.12) então

$$\|\gamma(x)\| = \|J_f(x)^T A_x \tilde{w}\| \leq \|J_f(x)\| \|A_x\| \|\tilde{w}\|,$$

e como A_x e \tilde{w} são limitados, e J_f é limitado sob compactos por ser contínua, $\gamma(x)$ é limitado por algum número real.

Se $x \in K$ é ponto crítico então $\gamma(x) = 0$.

Logo, $\gamma(K)$ é limitado.

Condição (iii.) Sejam $\{x_k\}_{k \in \mathbb{N}} \subseteq C$ não críticos e $x_k \rightarrow x \in C$ não crítico. Será mostrado que

$$\limsup_{k \rightarrow +\infty} J_f(x_k) \gamma(x_k) < 0.$$

Se $\gamma(x_k)$ é direção linear então $J_f(x_k) \gamma(x_k)$ já é limitada superiormente, na ordem parcial do \mathbb{R}^m , por valores negativos, pois a direção linear satisfaz a condição (iii.).

Pela continuidade de J_f e como $x_k \rightarrow x$ não crítico então $\exists N \in \mathbb{N}$ e $\exists L > 0$ tais que $\forall k > N$, $L < L_k$.

Por (4.12) da direção de Pareto descendente, se $\gamma(x_k) = -J_f(x)^T A_x \tilde{w}$ então $J_f(x)^T \gamma(x_k) \preceq -\beta L_x \mathbf{1}^m$. Logo,

$$\begin{aligned} J_f(x) \gamma(x_k) &\preceq -\beta L_x^2 \mathbf{1}^m \preceq -\beta L^2 \mathbf{1}^m < 0 \\ \implies \limsup_{k \rightarrow +\infty} J_f(x) \gamma(x_k) &\preceq -\beta L^2 \mathbf{1}^m < 0. \end{aligned}$$

Logo, a direção de Pareto descendente gera uma aplicação regular de direções. ■

Pela Proposição 22, o Algoritmo 2 com a direção de Pareto descendente satisfaz as condições do Teorema 1.

Um problema equivalente a (4.12) que também resulta na direção de Pareto descendente é descrito abaixo, que é uma versão mais conveniente para implementação.

Proposição 23. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciável em um aberto contendo C , $\beta > 0$ e $x \in C$*

Sejam $L_x \in \mathbb{R}$ e $A_x \in \mathbb{R}^{m \times m}$ matriz diagonal onde

$$L_x := \min_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|,$$

$$(A_x)_{jj} := \begin{cases} 0, & \text{se } L_x = 0, \\ \frac{L_x}{\|\nabla f_j(x)\|}, & j \in \{1, \dots, m\}, \text{ se } L_x > 0. \end{cases}$$

Sejam $\tilde{w} \in \mathbb{R}^n$ e $\tilde{\tau} \in \mathbb{R}$ definidas como

$$(\tilde{w}, \tilde{\tau}) = \begin{cases} \underset{\substack{(w, \tau) \in \mathbb{R}^m \times \mathbb{R} \\ x - J_f(x)^T w \in C}}{\text{argmin}} & \tau \\ \text{sujeito a} & J_f(x)J_f(x)^T A_x w + \tau \mathbf{1}^m \succeq 0, \\ & \mathbf{1}^T w \leq 1, \\ & w \succeq 0. \end{cases} \quad (4.13)$$

Seja $\tilde{v} \in \mathbb{R}^n$ definida como

$$\tilde{v} = \begin{cases} \text{direção linear}, & \text{se } \tilde{\tau} > -\beta L_x^2, \\ -J_f(x)^T \tilde{w}, & \text{se } \tilde{\tau} \leq -\beta L_x^2. \end{cases} \quad (4.14)$$

Então \tilde{v} é a direção de Pareto descendente em x .

Demonstração. Caso $L_x = 0$ então $\tilde{v} = 0 \in \mathbb{R}^n$ é solução de (4.13) e, portanto, direção central em x . Suponha $L_x \neq 0$.

Suponha $L_x \neq 0$. Note a equivalência

$$J_f(x)J_f(x)^T A_x w + \tau \mathbf{1}^m \succeq 0$$

$$\Longleftrightarrow \tau \geq \max_{j \in \{1, \dots, m\}} - \frac{L_x \nabla f_j(x)^T J_f(x)^T w}{\|\nabla f_j(x)\|}.$$

Por τ ser minimizado,

$$\tau = \max_{j \in \{1, \dots, m\}} - \frac{L_x \nabla f_j(x)^T J_f(x)^T w}{\|\nabla f_j(x)\|}.$$

Assim,

$$\tilde{w} = \begin{cases} \underset{\substack{w \in \mathbb{R}^m \\ x - J_f(x)^T A_x w \in C}}{\operatorname{argmin}} & \max_{j \in \{1, \dots, m\}} - \frac{L_x \nabla f_j(x)^T J_f(x)^T w}{\|\nabla f_j(x)\|} \\ \text{sujeito a} & \mathbf{1}^T w \leq 1, \\ & w \succeq 0, \end{cases}$$

que é o mesmo problema de (4.11) da Definição 14.

Se $\tilde{\tau} > -\beta L_x^2$ então $-J_f(x) J_f(x)^T A_x \tilde{w} \not\preceq -\beta L_x^2 \mathbf{1}^m$.

Se $\tilde{\tau} \leq -\beta L_x^2$ então $-J_f(x) J_f(x)^T A_x \tilde{w} \preceq -\beta L_x^2 \mathbf{1}^m$.

Portanto, (4.14) é o mesmo que (4.11) da Definição 14, logo, \tilde{v} é uma direção de Pareto descendente em x . ■

O problema (4.13) na Proposição 23 é um problema de programação linear, quando C pode ser definido por restrições lineares.

Pode-se comparar a direção linear com a direção de Pareto descendente quanto ao tempo necessário para computar. Ambas resolvem problemas de programação linear muito semelhantes.

A matriz que representa as restrições lineares na direção de Pareto descendente possui, no pior caso, $4m + 1$ linhas e $m + 1$ colunas, enquanto na direção linear, a matriz possui m linhas e $n + 1$ colunas. Normalmente, n é muito maior que m , e isso faz com que, em vários casos, o tempo necessário para computar a direção de Pareto descendente seja menor.

Porém, direção de Pareto descendente depende da direção linear pois nem sempre existe direção factível de descida que é combinação cônica dos gradientes negativos. Nestes casos, a direção de Pareto descendente leva mais tempo para computar que a direção linear.

A Figura 11 apresenta um exemplo com todas direções analisadas.

Exemplo com $f: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ e $C = \mathbb{R}^2$

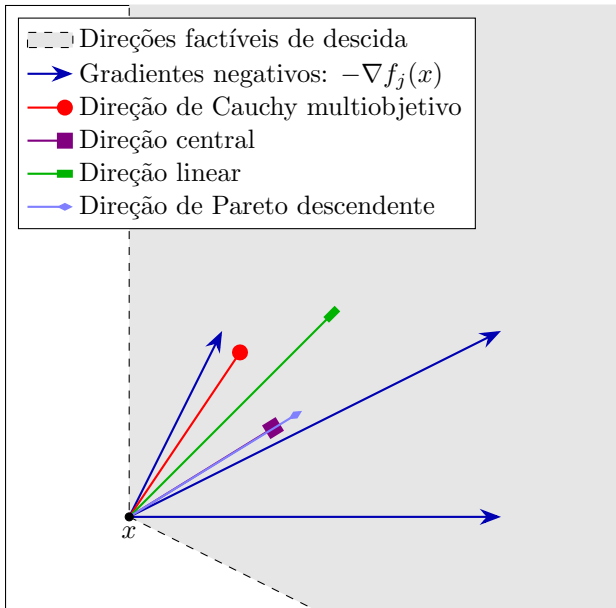


Figura 11 – Exemplo com todas direções factíveis de descida analisadas.

5 Experimentos numéricos

Todos os experimentos computacionais foram rodados no software *GNU Octave versão 4.2.2* (GNU, 2018), no sistema operacional *Ubuntu 18.04.1* (Canonical, 2018). Os códigos fontes gerados neste trabalho estão disponíveis em (GitHub, 2018).

Nos casos em que haviam apenas duas funções componentes e em que os gradientes negativos pertenciam ao conjunto viável $(C - x)$, utilizou-se as formas diretas da direção de Cauchy Multiobjetivo (ver Proposição 27 do Apêndice D) e da direção central (ver Proposição 28 do Apêndice D), a fim de se obter um melhor desempenho.

Para resolver os subproblemas de programação linear, da direção linear e Pareto descendente, foi utilizada a função `glpk` do Octave. Segundo a documentação da função, o problema de programação linear é resolvido pelo *revised simplex method* utilizando *dual-phase primal simplex*.

Para os subproblemas de programação quadrática, da direção de Cauchy multiobjetivo e central, foi usada a função `qp` do Octave. Segundo a documentação da função, o problema de programação quadrática é resolvido pelo algoritmo *non-negative least squares (NNLS)*.

O `glpk` apresentou mensagens de *instabilidade numérica* em alguns casos. Segundo a documentação, isso ocorre quando a solução não está mais dentro do conjunto viável devido a muitos erros de arredondamento acumulados. Valores maiores de Δ_x foram suficiente para eliminar a instabilidade numérica na maioria

dos casos. O parâmetro da direção linear usado foi

$$\Delta_x = 10^{-2} + \min_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|.$$

Com este Δ_x , as direções obtidas possuem norma proporcional à norma dos gradientes. Nos casos de instabilidade numérica, o termo 10^{-2} foi substituído por 10^{-1} , 10^0 , \dots , 10^3 , até uma solução ser obtida. Em caso ainda de instabilidade numérica utilizando o termo 10^3 , a rodada foi considerada como não convergente. Evitou-se usar inicialmente o termo constante grande, como 10^3 , pois isso gera direções com norma grande, e com isso pode ser necessário muitas iterações no *backtracking*, aumentando consideravelmente o tempo para convergência. Todos problemas que serão apresentados possuem o $\min_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|$ limitado sob compactos, o que faz com que Δ_x satisfaça todas condições necessárias da Definição 13.

O parâmetro da direção de Pareto descendente usado foi $\beta = 10^{-1}$. O parâmetro da condição de Armijo utilizado em todos os métodos foi $\sigma = 10^{-4}$. Os demais parâmetros foram especificados em cada problema.

Todos experimentos computacionais foram feitos considerando funções $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ onde $m \in \{2, 3\}$. Em funções com m maiores, a chance de existir alguma direção que forma um ângulo de mais de 90° com todos gradientes se torna menor, o que faz com que quase todos pontos sejam críticos.

5.1 Coleção de problemas

Nesta seção são apresentados diversos problemas para comparar os métodos analisados. Também foram utilizados na comparação resultados obtidos em (LUCAMBIO PÉREZ; PRU-

DENTE, 2018) com uma variação de um método chamado *Polak-Ribière-Polyak*, que é denotado no artigo por *PRP+*. Se trata de um método de gradientes conjugados não-linear para otimização vetorial.

O critério de parada para os métodos de direção de Cauchy multiobjetivo e central foi

$$\max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T v_k + \frac{1}{2} \|v_k\|^2 \geq -7,45 \cdot 10^{-8},$$

que é o mesmo critério usado para o *PRP+* em (LUCAMBIO PÉREZ; PRUDENTE, 2018). Assim como no artigo, foram permitidas no máximo 10.000 iterações.

Para os métodos de direção linear e Pareto descendente, o critério de parada foi

$$\max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T v_k \geq -7,45 \cdot 10^{-8}.$$

Cada problema foi rodado 20 vezes para cada método, com pontos iniciais aleatórios de um conjunto que é definido em cada problema, que coincide em alguns casos com o conjunto viável. Estes conjuntos foram os mesmos usados em (LUCAMBIO PÉREZ; PRUDENTE, 2018).

Os problemas foram divididos em três partes. Na Seção 5.1.1, problemas de funções convexas. Na Seção 5.1.2, problemas de funções não convexas. E na Seção 5.1.3, problemas que se tornam difíceis com n maiores.

5.1.1 Problemas de funções convexas

A Tabela 1 apresenta problemas convexas que são resolvidos, e os resultados são apresentados na Tabela 2. Os resultados para o método *PRP+* foram obtidos diretamente de (LUCAMBIO PÉREZ; PRUDENTE, 2018).

Pode-se observar que os métodos *Cauchy* e *central* obtiveram maior taxa de convergência que os métodos *linear* e *Pareto*. O método *central* obteve melhores resultados que o método *Cauchy* no problema *FDS*, mas nos demais problemas, os dois métodos obtiveram resultados semelhantes. De modo geral, o método *PRP+* obteve os melhores resultados.

O problema *FDS* foi obtido em (FLIEGE; DRUMMOND; SVAITER, 2009), o *Lov1* em (LOVISON, 2010) e os demais, *JOS1*, *MOP7* e *SP1*, em (HUBAND et al., 2006).

Problema	Função a ser minimizada
<p>FDS</p> <p>$x \in [-2, 2]^n$</p>	$f_1 : \frac{\sum_{k=1}^n k(x_k - k)^4}{n^2}$ $f_2 : \exp(\mathbf{1}^T x/n) + \ x\ ^2$ $f_3 : \frac{\sum_{k=1}^n k(n - k + 1) \exp(-x_k)}{n(n + 1)}$
<p>JOS1</p> <p>$x \in \mathbb{R}^n$</p>	$f_1 : \ x\ ^2/n$ $f_2 : \ x - 2(\mathbf{1}^n)\ ^2/n$
<p>Lov1</p> <p>$x \in \mathbb{R}^2$</p>	$f_1 : 1,05x_1^2 + 0,98x_2^2$ $f_2 : 0,99(x_1 - 3)^2 + 1,03(x_2 - 2,5)^2$
<p>MOP7</p> <p>$x \in [-400, 400]^2$</p>	$f_1 : \frac{(x_1 - 2)^2}{2} + \frac{(x_2 + 1)^2}{13} + 3$ $f_2 : \frac{(x_2 + x_1 - 3)^2}{36} + \frac{(x_2 - x_1 + 2)^2}{17} - 17$ $f_3 : \frac{(2x_2 + x_1 - 1)^2}{175} + \frac{(2x_2 - x_1)^2}{17} - 13$
<p>SP1</p> <p>$x \in \mathbb{R}^2$</p>	$f_1 : (x_1 - 1)^2 + (x_1 - x_2)^2$ $f_2 : (x_2 - 3)^2 + (x_1 - x_2)^2$

Tabela 1 – Problemas convexos de minimização.

		Cauchy	Central	Linear	Pareto	PRP
FDS	%	80	100	0	0	100
$n = 50$	It.	1984	61	-	-	46
$m = 3$	f	13492	61	-	-	507
$[-2, 2]^n$	T.	14,9s	0,5s	-	-	-
JOS1	%	100	100	5	85	100
$n = 1000$	It.	6877	6931	1716	4917	1
$m = 2$	f	6877	6931	3861	5048	18
$[-10^4, 10^4]^n$	T.	72,4s	39,8s	13,5s	24,8s	-
Lov1	%	100	100	100	100	100
$n = 2$	It.	4	6	20	14	3
$m = 2$	f	7	12	161	203	29
$[-100, 100]^n$	T.	0,0s	0,0s	0,0s	0,0s	-
MOP7	%	100	100	100	100	100
$n = 2$	It.	165	168	96	116	3
$m = 3$	f	165	168	100	117	52
$[-400, 400]^n$	T.	0,3s	0,3s	0,2s	0,2s	-
SP1	%	100	100	100	100	100
$n = 2$	It.	11	10	19	15	3
$m = 2$	f	27	24	156	139	28
$[-100, 100]^n$	T.	0,0s	0,0s	0,0s	0,0s	-

Tabela 2 – Taxa de sucesso e mediana do número de iterações, da quantidade de avaliações de f e do tempo, para cada método e problema. Os pontos iniciais foram tomados no conjunto indicado na primeira coluna da tabela. Os casos que não convergiram nos métodos de direção linear e Pareto descendente foram devidos à instabilidade numérica no `glpk`. No caso do método de Cauchy multiobjetivo, a causa foi máximo de iterações atingido. Não há informação de tempo para o *PRP+*.

5.1.2 Problemas de funções não convexas

A Tabela 4 apresenta resultados de problemas da Tabela 3.

Problema	Função a ser minimizada
FF1 $x \in \mathbb{R}^2$	$f_1 : 1 - \exp(-(x_1 - 1)^2 - (x_2 + 1)^2)$ $f_2 : 1 - \exp(-(x_1 + 1)^2 - (x_2 - 1)^2)$
Hil1 $x \in \mathbb{R}^2$	$f_1 : \cos(2\pi a(x)/360)b(x)$ $f_2 : \sin(2\pi a(x)/360)b(x)$ $a(x) = 45 + 40 \sin(2\pi x_1) + 25 \sin(2\pi x_2)$ $b(x) = 1 + 0,5 \cos(2\pi x_1)$
MMR5 $x \in [-5, 5]^n$	$f_1 : \left(\frac{1}{n} \sum_{k=1}^n b(x_k) \right)^{1/4}$ $f_2 : \left(\frac{1}{n} \sum_{k=1}^n b(x_k - 1,5) \right)^{1/4}$ $b(x_k) = x_k^2 - 10 \cos(2\pi x_k) + 10$
MOP5 $x \in [-30, 30]^2$	$f_1 : 0,5\ x\ ^2 + \sin(\ x\ ^2)$ $f_2 : \frac{(3x_1 - 2x_2 + 4)^2}{8} + \frac{(x_1 - x_2 + 1)^2}{27} + 15$ $f_3 : \frac{1}{\ x\ ^2 + 1} - 1,1 \exp(-\ x\ ^2)$
VU1 $x \in [-3, 3]^2$	$f_1 : \frac{1}{\ x\ ^2 + 1}$ $f_2 : x_1^2 + 3x_2^2 + 1$

Tabela 3 – Problemas não convexos de minimização.

O problema *Hil1* foi obtido em (HILLERMEIER, 2000), o *MMR5* em (MIGLIERINA; MOLHO; RECCHIONI, 2008) e os demais, *FF1*, *MOP5* e *VU1*, em (HUBAND et al., 2006).

Os métodos *Cauchy*, *central*, *linear* e *Pareto* obtiveram re-

sultados semelhantes de modo geral, com exceção do problema *MMR5* em que os métodos *linear* e *Pareto* convergiram em menos iterações. Vale notar também que o método *central* obteve melhores resultados que o método *PRP+* no problema *VU1*.

		Cauchy	Central	Linear	Pareto	PRP
FF1	%	100	100	100	100	100
$n = 2$	It.	27	42	21	25	13
$m = 2$	f	27	42	199	232	103
$[-1, 1]^n$	T.	0,0s	0,0s	0,0s	0,1s	-
Hil1	%	100	100	100	100	100
$n = 2$	It.	8	5	10	10	11
$m = 2$	f	25	18	163	169	95
$[0, 1]^n$	T.	0,0s	0,0s	0,0s	0,0s	-
MMR5	%	100	100	100	100	100
$n = 100$	It.	5557	5952	706	706	282
$m = 2$	f	5557	5952	7234	7234	1920
$[-5, 5]^n$	T.	7,3s	8,9s	1,8s	2,3s	-
MOP5	%	100	100	100	100	100
$n = 2$	It.	21	31	11	13	2
$m = 3$	f	21	31	36	56	21
$[-1, 1]^n$	T.	0,0s	0,1s	0,0s	0,0s	-
VU1	%	100	100	100	100	100
$n = 2$	It.	158	77	212	212	951
$m = 2$	f	384	77	1911	1911	4246
$[-3, 3]^n$	T.	0,1s	0,0s	0,4s	0,5s	-

Tabela 4 – Taxa de sucesso e mediana do número de iterações, da quantidade de avaliações de f e do tempo, para cada método e problema. Os pontos iniciais foram tomados no conjunto indicado na primeira coluna da tabela. Não há informação de tempo para o *PRP+*.

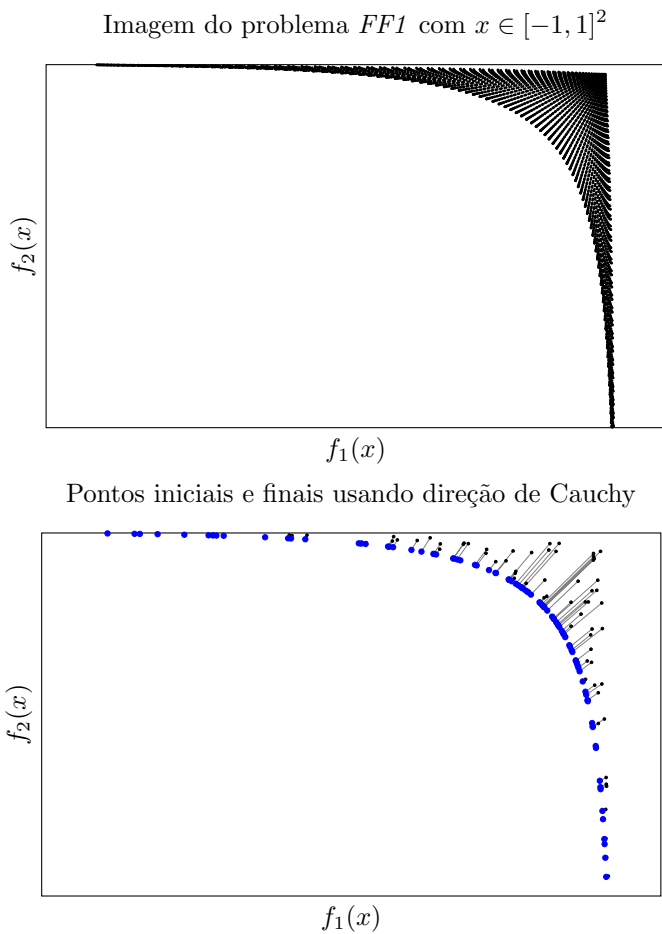
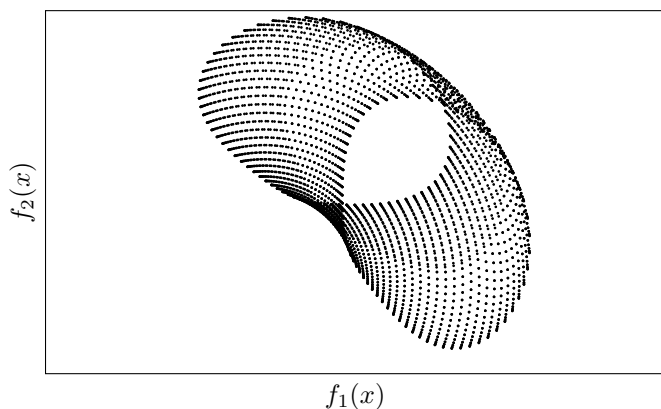


Figura 12 – Parede ótima obtida resolvendo 200 vezes o problema *FF1* pelo método da direção de Cauchy multiobjetivo. Cada ponto inicial (pontos menores) está ligado por uma linha à solução obtida (pontos maiores).

Imagem do problema *Hil1* com $x \in [0, 1]^2$



Pontos iniciais e finais resolvido com direção linear

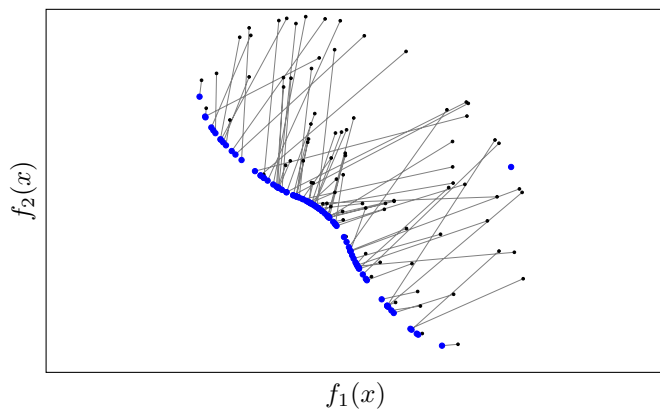
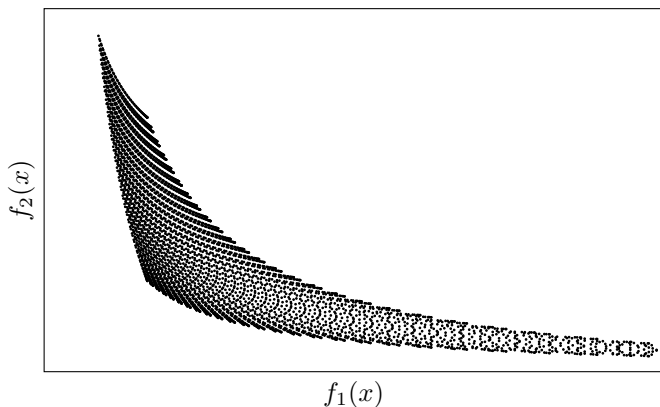


Figura 13 – Parede ótima obtida resolvendo 200 vezes o problema *Hil1* pelo método da direção linear. Cada ponto inicial (pontos menores) está ligado por uma linha à solução obtida (pontos maiores).

Imagem do problema $VU1$ com $x \in [-3, 3]^2$



Pontos iniciais e finais resolvido com direção central

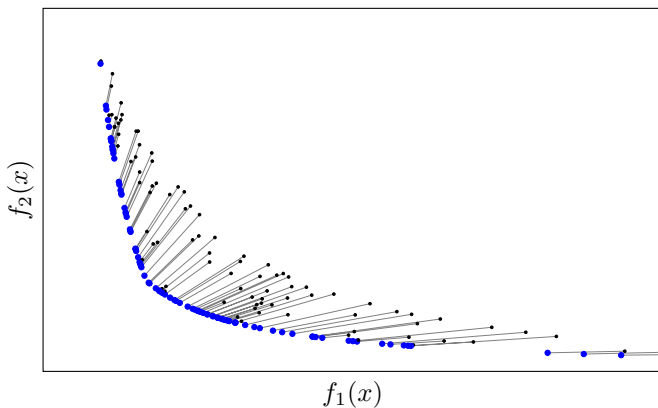


Figura 14 – Parede ótima obtida resolvendo 200 vezes o problema $VU1$ pelo método da direção central. Cada ponto inicial (pontos menores) está ligado por uma linha à solução obtida (pontos maiores).

As Figuras 12, 13 e 14 apresentam uma estimativa da parede

ótima dos problemas *FF1*, *Hil1* e *VU1* em 200 rodadas, onde em cada problema foi resolvido com um método diferente, utilizando como critério de parada $\max_{j \in \{1, \dots, m\}} \nabla f_j(x_k)^T v_k \geq -10^{-8}$.

5.1.3 Problemas difíceis

Os problemas *FDS* (Tabela 1) e *MMR5* (Tabela 3) foram resolvidos também para $n \geq 200$, onde estes problemas se tornam mais difíceis.

No caso do *FDS*, apenas o método da direção central convergiu para um ponto crítico. O método da direção linear, Cauchy e Pareto convergiram em menos de 10% das rodadas. Os resultados são apresentados na Tabela 5. Para $n > 2000$, o método não converge em menos de 15 minutos, que foi o tempo máximo permitido.

FDS usando direção central				
n	Sucesso	Iterações	Eval. f	Tempo
200	100%	67	67	2s
500	100%	68	68	15s
1000	100%	69	69	102s
2000	100%	72	72	883s

Tabela 5 – Problema FDS utilizando o método da direção central.

No caso do *MMR5*, apenas os métodos da direção linear e Pareto descendente convergiram para um ponto crítico. Os métodos Cauchy multiobjetivo e central atingiram o limite de iterações sem convergir. O desempenho do método da direção linear foi ligeiramente superior, e é apresentado na Tabela 6. Para $n > 2000$, o problema de programação linear apresentou muita instabilidade numérica no *glpk*.

MMR5 usando direção linear				
n	Sucesso	Iterações	Eval. f	Tempo
200	100%	730	6769	2s
500	100%	444	3105	2s
1000	100%	313	1627	3s
2000	100%	237	816	12s

Tabela 6 – Problema MMR5 utilizando o método da direção linear.

5.2 Portifólio

O problema de *portifólio* consiste em encontrar uma distribuição de *ativos* que traga o maior retorno com o menor risco possível.

Em um portfólio de $n \in \mathbb{N}$ ativos, a distribuição desejada é representada por um vetor $w \in \mathbb{R}^n$ onde $\mathbf{1}^T w = 1$ e $w \succeq 0$. Ou seja, o vetor w indica o peso que cada ativo tem dentro de um carteira de ativos.

Nesta aplicação, diversas bases de dados reais de portfólio foram utilizadas (BRUNI et al., 2016).

Para cada rodada, é gerada uma matriz $A \in \mathbb{R}^{52 \times n}$ contendo dados de ganhos e perdas de n ativos de um portfólio em um período de 52 semanas. O vetor $Aw \in \mathbb{R}^{52}$ corresponde ao retorno obtido em cada uma das 52 semanas, quando é usada a distribuição w de ativos. O retorno total negativo (para poder ser minimizado) é obtido como

$$f_1(w) = -\mathbf{1}^T Aw.$$

É gerada também uma matriz $C \in \mathbb{R}^{n \times n}$, que é a matriz de covariância entre os ativos, obtida da matriz A . O risco

associado pode ser definido como sendo a variância do portfólio (MARKOWITZ, 1952), que é dado por

$$f_2(w) = w^T C w.$$

Portanto, o problema a ser minimizado é

$$\begin{aligned} \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad & f(w) = \begin{bmatrix} f_1(w) \\ f_2(w) \end{bmatrix} = \begin{bmatrix} -\mathbf{1}^T A w \\ w^T C w \end{bmatrix} \\ \text{sujeito a} \quad & \mathbf{1}^T w = 1 \\ & w \succeq 0. \end{aligned}$$

Para verificar a qualidade da solução obtida, utiliza-se uma matriz $B \in \mathbb{R}^{12 \times n}$, que contém dados de ganhos e perdas das 12 semanas seguintes aos dados da matriz A . Portanto, verifica-se a qualidade do resultado calculando-se $\mathbf{1}^T B w$.

O artigo com as fontes de dados também apresenta os vetores $w \in [0, 1]^n$ para cada matriz A utilizando 6 modelos de seleção de portfólios diferentes, apresentados na Tabela 7. Para mais detalhes, ver (BRUNI et al., 2016).

Os portfólios usados estão indicados na Tabela 8 (BRUNI et al., 2016).

O critério de parada foi $\max_{j \in \{1, 2\}} \nabla f_j(w_k)^T v_k \geq -10^{-7}$, nos portfólios com menos de 100 ativos, e $\max_{j \in \{1, 2\}} \nabla f_j(w_k)^T v_k \geq -10^{-6}$ nos demais portfólios. Foram permitidas no máximo 10.000 iterações. O ponto inicial usado em cada rodada foi $x_0 = \frac{1}{n} \mathbf{1}^n$.

Os resultados obtidos estão na Tabela 9. Pode-se observar que os métodos analisados (direção Cauchy multiobjetivo, direção central, direção linear e direção de Pareto descendente) obtiveram

Nome	Descrição
CZeSD	Cumulative Zero-order epsilon Stochastic Dominance
RMZ_SSD	Roman-Mitra-Zviarovich Second-Order Stochastic Dominance
LR_ASSD	Lizyayev-Ruszczynski approximate Second-Order Stochastic Dominance
L_SSD	Luedtke Second-Order Stochastic Dominance
KP_SSD	Post-Kopa Second-Order Stochastic Dominance
MeanVar	Markowitz Mean-Variance

Tabela 7 – Modelos de seleção de portfólios usados para comparar com resultados obtidos usando os métodos de descida. Para detalhes destes modelos, ver ([BRUNI et al., 2016](#)).

Portfólio	Ativos (n)	Épocas
DowJones	28	110
NASDAQ100	82	46
FTSE100	83	56
SP500	442	46
NASDAQComp	1203	53
FF49Industries	49	190

Tabela 8 – Portfólios usados. A coluna *épocas* indica a quantidade de matrizes A e B diferentes que foram obtidas utilizando a base de dados do portfólio.

retornos intermediários em relação aos métodos específicos para problemas de portfólio.

A taxa de sucesso na convergência, o tempo médio e a quantidade média de iterações utilizadas são apresentadas na Tabela

Retorno trimestral médio em % ($1^T Bw$)						
Método	Dow Jones	NAS. 100	FTSE 100	SP 500	NAS. Comp	FF49 Indu.
Cauchy	2,5	3,7	2,8	2,2	1,9	5,1
Central	2,6	4,0	3,1	2,3	2,6	5,1
Linear	2,5	3,2	3,3	1,9	3,1	5,0
Pareto	2,4	3,2	3,3	1,9	3,2	5,1
CZeSD	2,7	4,7	2,5	3,1	3,6	6,0
KP_ssd	4,3	6,5	6,0	4,2	1,7	7,2
L_ssd	2,1	4,0	3,4	2,1	1,8	5,0
LR_assd	2,4	3,1	2,4	2,7	3,6	3,5
MeanVar	2,2	3,2	3,2	2,1	1,7	4,9
RMZ_ssd	2,0	3,3	4,0	2,5	4,1	5,4

Tabela 9 – Retorno médio, para cada portfólio e método. Os piores e melhores retornos estão destacados. Os resultados dos 6 modelos de seleção de portfólios foram obtidos diretamente de (BRUNI et al., 2016).

10. Pode-se observar que os métodos de programação linear (linear e Pareto descendente) chegaram a um ponto crítico mais rapidamente que os métodos de programação quadrática (Cauchy multiobjetivo e central), especialmente nos portfólios com maior quantidade de ativos.

Portifólio	Cauchy	Central	Linear	Pareto
Dow Jones	100% 236 it. 0,6s	100% 228 it. 0,6s	100% 25 it. 0,1s	100% 22 it. 0,1s
NASDAQ100	100% 274 it. 1,7s	100% 271 it. 1,5s	100% 28 it. 0,1s	100% 28 it. 0,1s
FTSE100	100% 222 it. 1,3s	100% 225 it. 1,3s	100% 23 it. 0,1s	100% 24 it. 0,1s
SP500	100% 35 it. 11,4s	100% 36 it. 10,1s	100% 17 it. 0,1s	100% 17 it. 0,1s
NASDAQComp	87% 16 it. 180,7s	91% 21 it. 207,4s	100% 42 it. 2,1s	100% 41 it. 2,1s
FF49Industries	100% 272 it. 1,0s	100% 244 it. 0,9s	100% 34 it. 0,1s	100% 32 it. 0,1s

Tabela 10 – Taxa de convergência, número de iterações e tempo médios para encontrar a solução em cada portfólio e método. As taxas de convergências inferiores a 100% foram devidas ao máximo de iterações atingidas no subproblema de programação quadrática.

Conclusão

Neste trabalho foram analisados os métodos de direção *Cauchy multiobjetivo*, *central*, *linear* e *Pareto descendente*. Os métodos de direção *Cauchy multiobjetivo* e *linear* já existiam na literatura. O método de direção de *Pareto descendente* é uma variação proposta de um método também já existente, e o método da direção *central* é uma variação proposta do método de direção de Cauchy multiobjetivo.

Nos Capítulos 2, 3 e 4 foi desenvolvida a teoria para mostrar que todo ponto limite gerado pelos métodos de descida analisados são ponto crítico ou ótimo de Pareto, sob certas condições na função $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ e no conjunto viável $C \subseteq \mathbb{R}^n$. Foi possível mostrar resultados mais fortes para os métodos *Cauchy multiobjetivo* e *central*, garantindo a convergência da sequência gerada sob certas hipóteses.

Este trabalho trouxe como contribuição a criação do conceito de *aplicação regular de direções*, que generaliza o conceito de *sequências gradient-related* de otimização escalar para o contexto de otimização multiobjetivo. Este conceito permite demonstrar mais facilmente a convergência das sequências geradas pelos métodos.

No Capítulo 5, uma coleção de problemas e uma aplicação de portfólio foram usadas para comparar os métodos analisados.

Na coleção de problemas convexos, os métodos *Cauchy multiobjetivo* e *central* apresentaram melhores resultados que os métodos *linear* e *Pareto descendente*. Em especial, o método da direção central conseguiu resolver o problema *FDS* para di-

mensões maiores, enquanto os três outros métodos não convergiram.

Na coleção de problemas não convexos, os métodos analisados apresentaram semelhantes, mas no problema *MMR5* com dimensão maiores foram os métodos *linear* e *Pareto descendente* que obtiveram melhores resultados.

Na aplicação de portfólio, os métodos *linear* e *Pareto descendente* se destacaram, especialmente nos portfólios com maior quantidade de ativos.

Concluiu-se, portanto, que cada método analisado pode se sobressair em certos problemas. A convergência demonstrada se comprovou na prática, mas houveram problemas de instabilidade numérica em alguns casos.

Existem muitas áreas de estudo possíveis no contexto deste trabalho, como analisar outras direções de descida: direção de Newton multiobjetivo, direções baseadas nos gradientes conjugados, entre outras.

Pode-se também aprofundar o estudo de problemas específicos de otimização multiobjetivo, como o problema de portfólio, que foi apresentado de forma apenas introdutória.

Outro área interessante é analisar as instabilidades numéricas, que foram presenciadas em alguns casos para as direções *linear* e *Pareto descendente*, de forma que se pudesse obter condições sob as quais se teria garantia de que a solução será encontrada de forma estável.

Referências

BERTSEKAS, D. *Convex analysis and optimization*. [S.l.]: Athena Scientific, 2003. ISBN 1-886529-45-0.

BERTSEKAS, D. P. *Nonlinear Programming*. Belmont, Massachussets: Athena Scientific, 1999. 791 p.

BRUNI, R. et al. Real-world datasets for portfolio selection and solutions of some stochastic dominance portfolio models. *Data in Brief*, Elsevier, v. 8, p. 858–862, 2016. ISSN 23523409. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352340916303997>.

Canonical. *Ubuntu*. 2018. Disponível em: <https://www.ubuntu.com/>.

CHIANG, T. C.; LIN, H. J. A simple and effective evolutionary algorithm for multiobjective flexible job shop scheduling. *International Journal of Production Economics*, Elsevier, v. 141, n. 1, p. 87–98, 2013. ISSN 09255273.

COMBETTES, P. L. Quasi-Fejérian Analysis of Some Optimization Algorithms. *Studies in Computational Mathematics*, n. 1, p. 1–38, 2001.

DENG, X. et al. Multi-objective optimization of cooling galleries inside pistons of a diesel engine. *Applied Thermal Engineering*, Elsevier Ltd, v. 132, p. 441–449, 2018. ISSN 13594311.

EBERLEIN, W. F. On Euler's Infinite Product for the Sine. *Journal of Mathematical Analysis And Applications*, v. 58, p. 147–151, 1977.

FLIEGE, J.; DRUMMOND, L. L.; SVAITER, B. B. Newton's method for multiobjective optimization. *SIAM Journal on Optimization*, v. 20, n. 2, p. 602–626, 2009. ISSN 10526234.

- FLIEGE, J.; SVAITER, B. F. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, v. 51, n. 3, p. 479–494, 2000. ISSN 14322994.
- FUKUDA, E. H.; DRUMMOND, L. M. G. a Survey on Multiobjective Descent Methods. *Pesquisa Operacional*, v. 34, n. 3, p. 585–620, 2014. ISSN 0101-7438.
- GitHub. *Códigos fontes do trabalho de conclusão de curso Otimização Multiobjetivo*. 2018. Disponível em: https://github.com/viniciusfazio/otimizacao_multiobjetivo.
- GNU. *GNU Octave*. 2018. Disponível em: <https://www.gnu.org/software/octave/>.
- GRAÑA DRUMMOND, L. M.; SVAITER, B. F. Métodos de Descida em Otimização Multiobjetivo. *30° Colóquio Brasileiro de Matemática*, IMPA, Rio de Janeiro, p. 49, 2015.
- HARADA, K. et al. Local Search for Multiobjective Function Optimization: Pareto Descent Method. *The 8th Annual Conference on Genetic and Evolutionary Computation*, p. 659–666, 2006. ISSN 1346-0714.
- HILLERMEIER, C. Generalized homotopy approach to multiobjective optimization. *Journal of Optimization Theory and Applications*, v. 110, n. 3, p. 557–583, 2000. ISSN 00223239.
- HUBAND, S. et al. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, v. 10, n. 5, p. 477–506, 2006. ISSN 1089778X.
- LOVISON, A. Singular Continuation: Generating Piece-wise Linear Approximations to Pareto Sets via Global Analysis. p. 1–21, 2010. ISSN 1052-6234.
- LUCAMBIO PÉREZ, L. R.; PRUDENTE, L. F. Nonlinear Conjugate Gradient Methods for Vector Optimization. *SIAM Journal on Optimization*, v. 28, n. 3, p. 2690–2720, 2018. ISSN 1052-6234.

LUCIANO, R. D. et al. Multi-objective optimization of cyclone separators in series based on computational fluid dynamics. *Powder Technology*, Elsevier B.V., v. 325, p. 452–466, 2018. ISSN 1873328X.

MARKOWITZ, H. Portfolio Selection. *The Journal of Finance*, v. 7, n. 1, p. 77, 1952. ISSN 15406261. Disponível em: <https://www.jstor.org/stable/2975974?origin=crossref>.

MARLER, R. T.; ARORA, J. S. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, v. 26, n. 6, p. 369–395, 2004. ISSN 1615147X.

MARTÍNEZ, J. M.; SANTOS, S. A. *Métodos computacionais de otimização*. IMECC - Unicamp, 1995. Disponível em: <http://www.ime.unicamp.br/{~}martinez/mslivro>.

MIETTINEN, K.; MÄKELÄ, M. M. On scalarizing functions in multiobjective optimization. *OR Spectrum*, v. 24, n. 2, p. 193–213, 2002. ISSN 01716468.

MIGLIERINA, E.; MOLHO, E.; RECCHIONI, M. C. Box-constrained multi-objective optimization: A gradient-like method without "a priori" scalarization. *European Journal of Operational Research*, v. 188, n. 3, p. 662–682, 2008. ISSN 03772217.

NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. [S.l.]: Springer, 1999. 634 p.

PRIFTIS, A. et al. Parametric design and multi-objective optimisation of containerships. *Ocean Engineering*, Elsevier Ltd, v. 156, n. October 2017, p. 347–357, 2018. ISSN 00298018.

QIN, H. et al. Multi-objective differential evolution with adaptive Cauchy mutation for short-term multi-objective optimal hydro-thermal scheduling. *Energy Conversion and Management*, Elsevier Ltd, v. 51, n. 4, p. 788–794, 2010. ISSN 01968904.

RIGONI, E.; POLES, S. NBI and MOGA-II, two complementary algorithms for Multi-Objective optimizations. *Practical Approaches to Multi-Objective Optimization*, 2005.

ZHOU, A. et al. Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, Elsevier B.V., v. 1, n. 1, p. 32–49, 2011. ISSN 22106502.

APÊNDICE A – Conceitos básicos

Definição 15. Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Então $\tilde{x} \in C \subseteq \mathbb{R}^n$ é dito *minimizador global* (ou *ótimo*) de f em C se

$$f(\tilde{x}) \leq f(x), \quad \forall x \in C.$$

Neste caso, C é dito ser o *conjunto viável*.

Proposição 24. *Sejam $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e $\tilde{x} \in \mathbb{R}^n$ o minimizador global de f .*

Então $-f(\tilde{x})$ é o valor máximo da função $-f$.

Demonstração.

$$f(\tilde{x}) \leq f(x) \implies -f(\tilde{x}) \geq -f(x), \quad \forall x \in \mathbb{R}^n.$$

■

Definição 16. Um conjunto $C \subseteq \mathbb{R}^n$ é dito *convexo* se

$$\lambda x + (1 - \lambda)y \in C, \quad \forall \lambda \in [0, 1], \forall x, y \in C.$$

Definição 17. Uma função $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é dita *sublinear* se todas as condições abaixo se aplicam:

i. $f(\lambda x) = \lambda f(x), \quad \forall x \in \mathbb{R}^n, \forall \lambda > 0.$

(homogênea positiva de grau 1)

Note que $f(0) = \lambda f(0), \forall \lambda > 0 \implies f(0) = 0.$

ii. $f(x + y) \leq f(x) + f(y), \quad \forall x, y \in \mathbb{R}^n.$

(subaditiva)

APÊNDICE B – Funções convexas

Definição 18. Se $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é *convexa* em $C \subseteq \mathbb{R}^n$ convexo então $\forall x, y \in C$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \quad \forall \lambda \in [0, 1].$$

No caso de desigualdade estrita, isto é, se $x \neq y$ e

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y), \quad \forall \lambda \in (0, 1),$$

então f é *estritamente convexa* em C .

Definição 19. Se $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ tem todas funções componentes convexas em $C \subseteq \mathbb{R}^n$ convexo então, $\forall x, y \in C$,

$$f(\lambda x + (1 - \lambda)y) \preceq \lambda f(x) + (1 - \lambda)f(y), \quad \forall \lambda \in [0, 1],$$

Neste caso, f será dita *convexa* em C . Equivalentemente,

$$f(y + \lambda(x - y)) \preceq f(y) + \lambda(f(x) - f(y)), \quad \forall \lambda \in [0, 1].$$

Se f tem ainda todas funções componentes estritamente convexas então, se $x \neq y$:

$$f(\lambda x + (1 - \lambda)y) \prec \lambda f(x) + (1 - \lambda)f(y), \quad \forall \lambda \in (0, 1),$$

e, neste caso, f será dita *estritamente convexa* em C .

Proposição 25. *Sejam $C \subseteq \mathbb{R}^n$ convexo, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável e convexa em C e $x \in C$. Então*

$$f(x + v) \geq f(x) + \nabla f(x)^T v, \quad \forall v \in \mathbb{R}^n, \quad x + v \in C.$$

Demonstração. Fixe $x, v \in C$. Como f é convexa, $\forall \lambda \in (0, 1)$:

$$\begin{aligned} f(x + \lambda v) &\leq f(x) + \lambda(f(x + v) - f(x)) \\ \implies \frac{f(x + \lambda v) - f(x)}{\lambda} + f(x) &\leq f(x + v) \\ \implies \lim_{\lambda \rightarrow 0^+} \frac{f(x + \lambda v) - f(x)}{\lambda} + f(x) &\leq f(x + v) \\ \implies \nabla f(x)^T v + f(x) &\leq f(x + v). \end{aligned}$$

■

Proposição 26. *Sejam $C \subseteq \mathbb{R}^n$ convexo e $f: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável e convexa em C . Então f é continuamente diferenciável em um aberto contendo C .*

Demonstração. Sejam $x \in C$, $v \in \mathbb{R}^n$ e $\{x_k\}_{k \in \mathbb{N}} \subseteq C$ uma sequência onde $x_k \rightarrow x$.

Pela definição de gradiente, vale que

$$\lim_{\lambda \rightarrow 0^+} \frac{f(x + \lambda v) - f(x)}{\lambda} = \nabla f(x)^T v. \quad (\text{B.1})$$

Por outro lado, pela Proposição 25,

$$\begin{aligned} f(x + \lambda v) &\geq f(x) + \nabla f(x)^T \lambda v, \quad \forall \lambda \in (0, 1) \\ \implies \frac{f(x + \lambda v) - f(x)}{\lambda} &\geq \nabla f(x)^T v, \quad \forall \lambda \in (0, 1). \end{aligned} \quad (\text{B.2})$$

Assim, usando (B.1) e (B.2), conclui-se que

$$\inf_{\lambda > 0} \frac{f(x + \lambda v) - f(x)}{\lambda} = \nabla f(x)^T v. \quad (\text{B.3})$$

Seja $\epsilon > 0$. Por (B.3), $\exists \lambda \in (0, 1)$ tal que

$$\frac{f(x + \lambda v) - f(x)}{\lambda} < \nabla f(x)^T v + \epsilon. \quad (\text{B.4})$$

Novamente pela Proposição 25, agora com x_k ,

$$\nabla f(x_k)^T v \leq \frac{f(x_k + \lambda v) - f(x_k)}{\lambda}. \quad (\text{B.5})$$

Como f é contínua,

$$\lim_{k \rightarrow +\infty} \frac{f(x_k + \lambda v) - f(x_k)}{\lambda} = \frac{f(x + \lambda v) - f(x)}{\lambda}. \quad (\text{B.6})$$

Tomando \limsup em (B.5) e usando (B.4) e (B.6),

$$\limsup_{k \rightarrow +\infty} \nabla f(x_k)^T v \leq \frac{f(x + \lambda v) - f(x)}{\lambda} < \nabla f(x)^T v + \epsilon, \quad (\text{B.7})$$

e fazendo $\epsilon \rightarrow 0$,

$$\limsup_{k \rightarrow +\infty} \nabla f(x_k)^T v \leq \nabla f(x)^T v.$$

Repetindo (B.7) com $-v$, pode-se inferir

$$\begin{aligned} \limsup_{k \rightarrow +\infty} \nabla f(x_k)^T (-v) &< \nabla f(x)^T (-v) + \epsilon \\ \implies \liminf_{k \rightarrow +\infty} \nabla f(x_k)^T v &> \nabla f(x)^T v - \epsilon, \end{aligned}$$

e fazendo $\epsilon \rightarrow 0$,

$$\liminf_{k \rightarrow +\infty} \nabla f(x_k)^T v \geq \nabla f(x)^T v.$$

Assim,

$$\begin{aligned} \limsup_{k \rightarrow +\infty} \nabla f(x_k)^T v &\leq \nabla f(x)^T v \leq \liminf_{k \rightarrow +\infty} \nabla f(x_k)^T v \\ \implies \lim_{k \rightarrow +\infty} \nabla f(x_k)^T v &= \nabla f(x)^T v. \end{aligned}$$

Como a sequência $\{x_k\}$ e v foram arbitrários, isso implica que ∇f é contínua em C , o que faz com que f seja continuamente diferenciável em um aberto contendo C . ■

APÊNDICE C – Exemplos adicionais

Exemplo 1 (Sequência não convergente). *Seja $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definida como o paraboloide*

$$f(x) = \frac{1}{2}\|x\|^2.$$

Sejam $C = \mathbb{R}^2$ o conjunto viável, $\sigma \in (0, 1)$ e $x_0 \in \mathbb{R}^2 \setminus \{0\}$.

Para a iteração $k \in \mathbb{N}$ e no ponto x_k , considere as seguintes variáveis para o Algoritmo 1:

$$\begin{cases} v_k = \begin{bmatrix} -\operatorname{sen} \alpha_k & -\cos \alpha_k \\ \cos \alpha_k & -\operatorname{sen} \alpha_k \end{bmatrix} x_k, & \alpha_k = \frac{1}{k+1}, \\ t_k = 2(1 - \sigma) \operatorname{sen} \alpha_k. \end{cases}$$

Então v_k é direção factível de descida e t_k satisfaz a condição do [Passo 2]. Além disso,

$$\lim_{k \rightarrow +\infty} \|x_k\| \geq \frac{2}{\pi} \|x_0\|,$$

de onde se conclui que todo ponto limite da sequência gerada $\{x_k\}$ não é um ponto crítico.

Demonstração. Seja $k \in \mathbb{N}$. Note que

$$0 < \alpha_k < 1 \quad \wedge \quad \nabla f(x_k) = x_k.$$

Primeiramente será mostrado que v_k é direção factível de descida.

Seja $x_k = \begin{bmatrix} a \\ b \end{bmatrix}$, com $a, b \in \mathbb{R}$.

$$\text{Então } f(x_k) = \frac{\|x_k\|^2}{2} = \frac{a^2 + b^2}{2} \text{ e}$$

$$\begin{aligned} \nabla f(x_k)^T v_k &= x_k^T \begin{bmatrix} -\sin \alpha_k & -\cos \alpha_k \\ \cos \alpha_k & -\sin \alpha_k \end{bmatrix} x_k \\ &= \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} -a \sin \alpha_k & -b \cos \alpha_k \\ a \cos \alpha_k & -b \sin \alpha_k \end{bmatrix} \\ &= -ab \cos \alpha_k - a^2 \sin \alpha_k - b^2 \sin \alpha_k + ab \cos \alpha_k \\ &= -(a^2 + b^2) \sin \alpha_k \\ &= -\|x_k\|^2 \sin \alpha_k \\ \implies \nabla f(x_k)^T v_k &= -\|x_k\|^2 \sin \alpha_k \end{aligned} \tag{C.1}$$

Como $\alpha_k > 0$ então

$$\begin{aligned} \nabla f(x_k)^T v_k &= -\|x_k\|^2 \sin \alpha_k < 0 \\ \iff -\|x_k\|^2 &< 0 \\ \iff \|x_k\| &> 0. \end{aligned}$$

Ou seja, o único ponto crítico de f é o vetor nulo. E se x_k não for ponto crítico então, por (C.1), v_k é direção factível de descida.

Agora será demonstrado que t_k satisfaz a condição do [Passo 2]. Antes, note que

$$\begin{aligned} 2f(x_k + t_k v_k) &= (a(1 - t_k \sin \alpha_k) - t_k b \cos \alpha_k)^2 \\ &\quad + (b(1 - t_k \sin \alpha_k) + t_k a \cos \alpha_k)^2 \\ &= a^2(1 - t_k \sin \alpha_k)^2 + t_k^2 b^2 \cos^2 \alpha_k \\ &\quad + b^2(1 - t_k \sin \alpha_k)^2 + t_k^2 a^2 \cos^2 \alpha_k \end{aligned}$$

$$\begin{aligned}
& -2ab(1 - t_k \sin \alpha_k)t_k \cos \alpha_k \\
& + 2ab(1 - t_k \sin \alpha_k)t_k \cos \alpha_k \\
& = (a^2 + b^2)((1 - t_k \sin \alpha_k)^2 + t_k^2 \cos^2 \alpha_k) \\
& = \|x_k\|^2(1 - 2t_k \sin \alpha_k + t_k^2) \\
& = \|x_k\|^2(1 + t_k^2) - 2t_k\|x_k\|^2 \sin \alpha_k,
\end{aligned}$$

que juntando com (C.1), chega-se em

$$\begin{aligned}
2f(x_k + t_k v_k) &= \|x_k\|^2(1 + t_k^2) + 2t_k \nabla f(x_k)^T v_k \\
\implies f(x_k + t_k v_k) &= f(x_k) + t_k^2 f(x_k) + t_k \nabla f(x_k)^T v_k. \quad (\text{C.2})
\end{aligned}$$

Logo, para garantir um t_k adequado, a condição do [Passo 2] deve ser satisfeita:

$$f(x_k + t_k v_k) \leq f(x_k) + \sigma t_k \nabla f(x_k)^T v_k.$$

Note que só faz sentido definir t_k se x_k não for ponto crítico. O que implica, neste exemplo, que $\|x_k\| > 0$. Substituindo $f(x_k + t_k v_k)$ de (C.2) na condição do [Passo 2] chega-se em:

$$\begin{aligned}
f(x_k) + t_k^2 f(x_k) + t_k \nabla f(x_k)^T v_k &\leq f(x_k) + \sigma t_k \nabla f(x_k)^T v_k \\
\iff t_k^2 f(x_k) &\leq (\sigma - 1)t_k \nabla f(x_k)^T v_k \\
\iff t_k \frac{\|x_k\|^2}{2} &\leq (\sigma - 1) \nabla f(x_k)^T v_k \\
\iff t_k &\leq \frac{2(\sigma - 1) \nabla f(x_k)^T v_k}{\|x_k\|^2} \\
\iff t_k &\leq \frac{2(1 - \sigma) \|x_k\|^2 \sin \alpha_k}{\|x_k\|^2} \\
\iff 0 < t_k &\leq 2(1 - \sigma) \sin \alpha_k.
\end{aligned}$$

Então, escolhendo $t_k = 2(1 - \sigma) \sin \alpha_k$, fica satisfeita a condição do [Passo 2].

Será mostrado agora que cada iteração decresce uma quantidade insuficiente a norma do ponto da iteração anterior para que o ponto crítico deste exemplo seja atingido. Supondo que x_k não é um ponto crítico (neste caso, significa que $\|x_k\| > 0$) e utilizando (C.2), pode-se deduzir que:

$$\begin{aligned}
 \|x_{k+1}\| &= \|x_k + t_k v_k\| \\
 &= \sqrt{2f(x_k + t_k v_k)} \\
 &= \sqrt{2f(x_k)(1 - 2t_k \sin \alpha_k + t_k^2)} \\
 &= \|x_k\| \sqrt{1 - 2t_k \sin \alpha_k + t_k^2} \\
 \implies \frac{\|x_{k+1}\|}{\|x_k\|} &= \sqrt{1 - 2t_k \sin \alpha_k + t_k^2}.
 \end{aligned}$$

Analisando a razão, e substituindo o t_k ,

$$\begin{aligned}
 \frac{\|x_{k+1}\|}{\|x_k\|} &= \sqrt{1 - 2t_k \sin \alpha_k + t_k^2} \\
 &= \sqrt{1 - 2(2(1 - \sigma) \sin \alpha_k) \sin \alpha_k + (2(1 - \sigma) \sin \alpha_k)^2} \\
 &= \sqrt{1 - 4(1 - \sigma) \sin^2 \alpha_k + 4(1 - \sigma)^2 \sin^2 \alpha_k} \\
 &= \sqrt{1 + 4\sigma \sin^2 \alpha_k - 8\sigma \sin^2 \alpha_k + 4\sigma^2 \sin^2 \alpha_k} \\
 &= \sqrt{1 - 4\sigma \sin^2 \alpha_k + 4\sigma^2 \sin^2 \alpha_k} \\
 &= \sqrt{1 + 4\sigma(\sigma - 1) \sin^2 \alpha_k}.
 \end{aligned}$$

Dentro da raiz existe um polinômio de segundo grau na variável σ , com mínimo quando $\sigma = \frac{1}{2}$, que é o caso em que a norma de x_k mais decresce. Então, para obter o maior decréscimo da norma de x_k , será fixado $\sigma = \frac{1}{2}$:

$$\begin{aligned}
 \frac{\|x_{k+1}\|}{\|x_k\|} &= \sqrt{1 + \sin^2 \alpha_k} = \sqrt{1 + \sin^2 \frac{1}{k+1}} \\
 \implies \|x_{k+1}\| &= \|x_k\| \sqrt{1 + \sin^2 \frac{1}{k+1}}.
 \end{aligned}$$

Com esta relação, pode-se determinar a norma com $k \rightarrow +\infty$:

$$\lim_{k \rightarrow +\infty} \|x_k\| = \|x_0\| \prod_{k=1}^{+\infty} \sqrt{1 - \frac{\sin^2 \frac{1}{k+1}}{k+1}} = \|x_0\| \prod_{k=1}^{+\infty} \cos \frac{1}{k+1}. \quad (\text{C.3})$$

É possível mostrar que $\prod_{k=1}^{+\infty} \cos \frac{1}{k+1} > 0$. Para isso, será usado o produto infinito de Euler para o seno (EBERLEIN, 1977), que é dado pela identidade:

$$\sin \alpha = \alpha \prod_{k=1}^{+\infty} \left(1 - \frac{\alpha^2}{k^2 \pi^2}\right), \quad \forall \alpha \in \mathbb{R}.$$

Em particular,

$$1 = \sin \frac{\pi}{2} = \frac{\pi}{2} \prod_{k=1}^{+\infty} \left(1 - \frac{\frac{\pi^2}{2^2}}{k^2 \pi^2}\right) \Rightarrow \frac{2}{\pi} = \prod_{k=1}^{+\infty} \left(1 - \frac{1}{4k^2}\right).$$

Sabendo que a expansão em Taylor do cosseno é

$$\cos \frac{1}{k} = 1 - \frac{1}{4k^2} + \frac{1}{24k^4} - \frac{1}{720k^6} + \dots > 1 - \frac{1}{4k^2},$$

chega-se à inequação

$$0 < \frac{2}{\pi} = \prod_{k=1}^{+\infty} \left(1 - \frac{1}{4k^2}\right) < \prod_{k=1}^{+\infty} \cos \frac{1}{k} < \prod_{k=1}^{+\infty} \cos \frac{1}{k+1} < 1. \quad (\text{C.4})$$

Utilizando (C.4) em (C.3), chega-se que

$$\lim_{k \rightarrow +\infty} \|x_k\| = \|x_0\| \prod_{k=1}^{+\infty} \cos \frac{1}{k+1} > \|x_0\| \frac{2}{\pi}.$$

Portanto, se $\|x_0\| \neq 0$ então nenhum ponto limite de $\{x_k\}$ converge para um ponto crítico pois $\liminf_{k \in \mathbb{N}} \|x_k\| > 0$. A Figura 15 ilustra essas iterações. ■

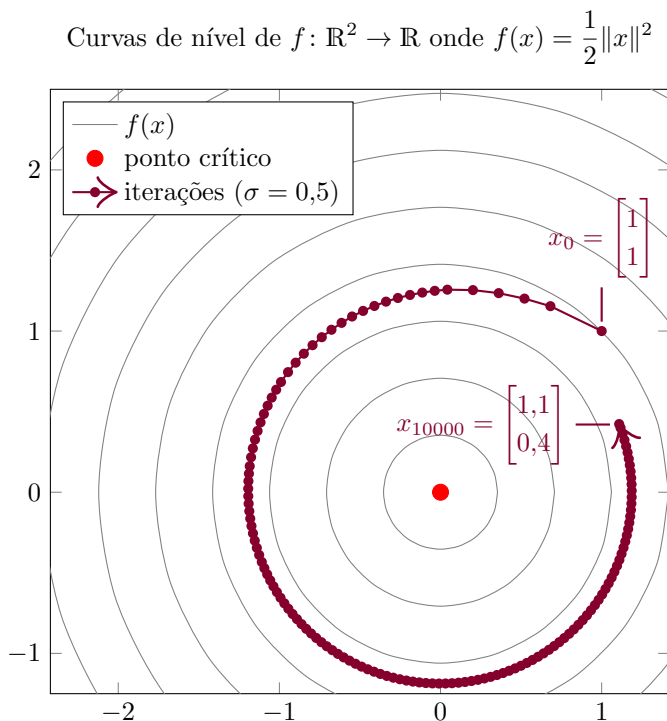


Figura 15 – Gráfico de $f(x)$ com 10.000 iterações do método de descida (ver Exemplo 1). Fica claro que a norma do ponto iterado só se reduziu nas primeiras iterações, e após isso os pontos apenas circulam ao redor do ponto crítico mas sem se aproximar, conforme previsto e demonstrado.

APÊNDICE D – Formas diretas

Os problemas que precisam ser resolvidos para encontrar a direção de Cauchy multiobjetivo e a direção central são problemas de programação quadrática, que nem sempre podem ser computados rapidamente. No caso particular em que $m = 2$ e em que todos gradientes negativos estão no conjunto viável $(C - x)$, é possível calcular essas direções de forma direta, envolvendo apenas algumas multiplicações entre vetores, como é apresentado nas Proposições 27 e 28.

Proposição 27. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^2$ diferenciável em um aberto contendo C e $x \in C$. Suponha $x - \nabla f_1(x) \in C$ e $x - \nabla f_2(x) \in C$.*

Seja

$$\tilde{\lambda} = \begin{cases} 0, & \nabla f_1(x) = \nabla f_2(x), \\ \frac{\nabla f_1(x)^T (\nabla f_1(x) - \nabla f_2(x))}{\|\nabla f_1(x) - \nabla f_2(x)\|^2}, & \nabla f_1(x) \neq \nabla f_2(x). \end{cases}$$

$$\tilde{v} = \begin{cases} -\nabla f_1(x), & \text{se } \tilde{\lambda} < 0, \\ -\nabla f_2(x), & \text{se } \tilde{\lambda} > 1, \\ (\lambda - 1)\nabla f_1(x) - \lambda\nabla f_2(x), & \text{se } \tilde{\lambda} \in [0, 1]. \end{cases}$$

Então \tilde{v} é direção de Cauchy multiobjetivo.

Demonstração. Como $x - \nabla f_1(x) \in C$ e $x - \nabla f_2(x) \in C$ por hipótese então, pela convexidade de C ,

$$\lambda(x - \nabla f_2(x)) + (1 - \lambda)(x - \nabla f_1(x)) \in C$$

$$\implies x + (\lambda - 1)\nabla f_1(x) - \lambda\nabla f_2(x) \in C, \quad \forall \lambda \in [0, 1]. \quad (\text{D.1})$$

Pela Proposição 15, $\exists \tilde{\lambda} \in [0, 1]$ tal que a direção de Cauchy multiobjetivo é definida como

$$\tilde{v} = \Pi_C(x + (\tilde{\lambda} - 1)\nabla f_1(x) - \tilde{\lambda}\nabla f_2(x)) - x,$$

onde Π_C é a projeção em C , definida na Proposição 8.

Por (D.1), $x + (\tilde{\lambda} - 1)\nabla f_1(x) - \tilde{\lambda}\nabla f_2(x) \in C$, logo,

$$\tilde{v} = (\tilde{\lambda} - 1)\nabla f_1(x) - \tilde{\lambda}\nabla f_2(x). \quad (\text{D.2})$$

Defina $(C - x) := \{v \in \mathbb{R}^n : x + v \in C\}$.

Para facilitar a notação, defina $\psi: [0, 1] \rightarrow \mathbb{R}^n$ como

$$\begin{aligned} \psi(\lambda) &= (\lambda - 1)\nabla f_1(x) - \lambda\nabla f_2(x) \\ \implies \psi'(\lambda) &= \nabla f_1(x) - \nabla f_2(x). \end{aligned}$$

Pela definição de direção de Cauchy multiobjetivo, e por (D.2), pode-se deduzir que

$$\tilde{\lambda} = \operatorname{argmin}_{\lambda \in [0, 1]} \max_{j \in \{1, 2\}} \nabla f_j(x)^T \psi(\lambda) + \frac{1}{2} \|\psi(\lambda)\|^2,$$

é tal que $\tilde{v} = \psi(\tilde{\lambda})$.

A ideia é usar condições de otimalidade do Teorema 3.

Defina $g: [0, 1] \rightarrow \mathbb{R}^2$ e $h: [0, 1] \rightarrow \mathbb{R}$ como

$$\begin{aligned} g_j(\lambda) &= \nabla f_j(x)^T \psi(\lambda) + \frac{1}{2} \|\psi(\lambda)\|^2, \quad \forall j \in \{1, 2\}, \\ \implies g'_j(\lambda) &= \nabla f_j(x)^T \psi'(\lambda) + \psi(\lambda)^T \psi'(\lambda), \\ h(\lambda) &= \max_{j \in \{1, 2\}} g_j(\lambda). \end{aligned}$$

Então, o minimizador de h em $[0, 1]$ é $\tilde{\lambda}$. Se $\nabla f_1(x) = \nabla f_2(x)$ então h é constante, portanto, qualquer $\tilde{\lambda} \in [0, 1]$ gera a direção de Cauchy multiobjetivo, por exemplo, $\tilde{v} = \psi(0)$.

Suponha $\nabla f_1(x) \neq \nabla f_2(x)$.

Como g é convexa e diferenciável, $[0, 1]$ é convexo, e h é o máximo das funções componentes de g , então as condições do Teorema 3 estão satisfeitas. Esse teorema garante que se \tilde{v} minimiza h então $\exists \mu \in [0, 1]$ tal que duas condições são satisfeitas. Uma das condições é

$$\begin{aligned}
 & \mu g_1(\tilde{\lambda}) + (1 - \mu)g_2(\tilde{\lambda}) = \max_{j \in \{1, 2\}} g_j(\tilde{\lambda}) \\
 \iff & \mu \nabla f_1(x)^T \psi(\tilde{\lambda}) + (1 - \mu) \nabla f_2(x)^T \psi(\tilde{\lambda}) + \frac{1}{2} \|\psi(\tilde{\lambda})\|^2 \\
 & = \max_{j \in \{1, 2\}} \nabla f_j(x)^T \psi(\tilde{\lambda}) + \frac{1}{2} \|\psi(\tilde{\lambda})\|^2 \\
 \iff & \nabla f_2(x)^T \psi(\tilde{\lambda}) + \mu \psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = \max_{j \in \{1, \dots, 2\}} \nabla f_j(x)^T \psi(\tilde{\lambda}) \\
 \iff & \mu \psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = \max\{0, (\nabla f_1(x) - \nabla f_2(x))^T \psi(\tilde{\lambda})\} \\
 \iff & \mu \psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = \max\{0, \psi'(\tilde{\lambda})^T \psi(\tilde{\lambda})\}.
 \end{aligned}$$

Portanto, se $\psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) < 0$ então $\mu = 0$.

Se $\psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) > 0$ então $\mu = 1$.

Se $\psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = 0$ então $\mu \in [0, 1]$.

A outra condição do Teorema 3 é que, $\forall \lambda \in [0, 1]$,

$$\begin{aligned}
 & \left(\mu g'_1(\tilde{\lambda}) + (1 - \mu)g'_2(\tilde{\lambda}) \right) (\lambda - \tilde{\lambda}) \geq 0 \\
 \iff & \left(g'_2(\tilde{\lambda}) + \mu(g'_1(\tilde{\lambda}) - g'_2(\tilde{\lambda})) \right) (\lambda - \tilde{\lambda}) \geq 0 \\
 \iff & \left(\nabla f_2(x)^T \psi'(\tilde{\lambda}) + \psi(\tilde{\lambda})^T \psi'(\tilde{\lambda}) + \mu \|\psi'(\tilde{\lambda})\|^2 \right) (\lambda - \tilde{\lambda}) \geq 0
 \end{aligned} \tag{D.3}$$

Como visto, se $\psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = 0$ então qualquer $\mu \in [0, 1]$ pode ser tomado. Portanto, para isso ocorrer,

$$\psi(\tilde{\lambda})^T \psi'(\tilde{\lambda}) = 0$$

$$\begin{aligned} &\Longleftrightarrow (-\nabla f_1(x) + \tilde{\lambda}\psi'(\tilde{\lambda}))^T \psi'(\tilde{\lambda}) = 0 \\ &\Longleftrightarrow \tilde{\lambda} = \frac{\nabla f_1(x)^T \psi'(\tilde{\lambda})}{\|\psi'(\tilde{\lambda})\|^2}. \end{aligned}$$

Logo, se $\tilde{\lambda} = \frac{\nabla f_1(x)^T \psi'(\lambda)}{\|\psi'(\lambda)\|^2} \in [0, 1]$ então $\tilde{\lambda}$ é solução do problema.

Para o caso $\frac{\nabla f_1(x)^T \psi'(\lambda)}{\|\psi'(\lambda)\|^2} > 1$, note que $\psi(\lambda)^T \psi'(\lambda)$ depende linearmente de λ com multiplicador positivo ($\|\psi'(\lambda)\|^2$). Portanto, para λ menor, $\psi(\lambda)^T \psi'(\lambda)$ será menor, o que implica que, neste caso, $\psi(\lambda)^T \psi'(\lambda)$ será sempre negativa no intervalo $[0, 1]$.

Dessa forma, $\mu = 0$. Então, é necessário escolher $\tilde{\lambda}$ em (D.3) tal que,

$$\left(\nabla f_2(x)^T \psi'(\tilde{\lambda}) + \psi(\tilde{\lambda})^T \psi'(\tilde{\lambda}) \right) (\lambda - \tilde{\lambda}) \geq 0$$

Pode-se tomar $\tilde{\lambda} = 1$ neste caso, o que resulta em

$$\left(\nabla f_2(x)^T \psi'(1) - \nabla f_2(x)^T \psi'(1) \right) (\lambda - 1) = 0 \geq 0.$$

De forma análoga, para o caso $\frac{\nabla f_1(x)^T \psi'(\lambda)}{\|\psi'(\lambda)\|^2} < 0$ a função $\psi(\lambda)^T \psi'(\lambda)$ é sempre positiva no intervalo $[0, 1]$. Portanto, $\mu = 1$. Então, é necessário escolher $\tilde{\lambda}$ em (D.3) tal que

$$\left(\nabla f_2(x)^T \psi'(\tilde{\lambda}) + \psi(\tilde{\lambda})^T \psi'(\tilde{\lambda}) + \|\psi'(\tilde{\lambda})\|^2 \right) (\lambda - \tilde{\lambda}) \geq 0.$$

Pode-se tomar $\tilde{\lambda} = 0$ neste caso, o que resulta em

$$\begin{aligned} &\left(\nabla f_2(x)^T \psi'(0) - \nabla f_1(x)^T \psi'(0) + \|\psi'(0)\|^2 \right) \lambda \\ &= (-\|\psi'(0)\|^2 + \|\psi'(0)\|^2) \lambda = 0 \geq 0. \end{aligned}$$

■

Proposição 28. *Suponha o conjunto viável $C \subseteq \mathbb{R}^n$ não vazio, convexo e fechado, $f: \mathbb{R}^n \rightarrow \mathbb{R}^2$ diferenciável em um aberto contendo C e $x \in C$. Defina $L_x \in \mathbb{R}$ como*

$$L_x := \min_{j \in \{1, \dots, m\}} \|\nabla f_j(x)\|.$$

Suponha $L_x > 0$ e que

$$x - \frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} \in C \quad \wedge \quad x - \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|} \in C.$$

Então $\tilde{v} = \frac{-L_x}{2} \left(\frac{\nabla f_1(x)}{\|\nabla f_1(x)\|} + \frac{\nabla f_2(x)}{\|\nabla f_2(x)\|} \right)$ é direção central.

Demonstração. Esta demonstração é bem parecida com a Proposição 27.

Pela convexidade de C ,

$$x + (\lambda - 1) \frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} - \lambda \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|} \in C, \quad \forall \lambda \in [0, 1]. \quad (\text{D.4})$$

Pela demonstração da Proposição 18, $\exists \tilde{\lambda} \in [0, 1]$ tal que a direção linear é definida como

$$\tilde{v} = \Pi_C \left(x + (\tilde{\lambda} - 1) \frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} - \tilde{\lambda} \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|} \right) - x,$$

onde Π_C é a projeção em C , da Proposição 8. Utilizando (D.4),

$$\tilde{v} = (\tilde{\lambda} - 1) \frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} - \tilde{\lambda} \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|}. \quad (\text{D.5})$$

Defina $(C - x) := \{v \in \mathbb{R}^n : x + v \in C\}$.

Para facilitar a notação, defina $\psi: [0, 1] \rightarrow \mathbb{R}^n$ como

$$\psi(\lambda) = (\lambda - 1) \frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} - \lambda \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|}$$

$$\implies \psi'(\lambda) = \frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} - \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|}.$$

Pela definição de direção central, e por (D.5), pode-se deduzir que

$$\tilde{\lambda} = \operatorname{argmin}_{\lambda \in [0,1]} \max_{j \in \{1,2\}} \frac{L_x \nabla f_j(x)}{\|\nabla f_j(x)\|} \psi(\lambda) + \frac{1}{2} \|\psi(\lambda)\|^2,$$

é tal que $\tilde{v} = \psi(\tilde{\lambda})$.

A ideia é usar condições de otimalidade do Teorema 3.

Defina $g: [0, 1] \rightarrow \mathbb{R}^2$ e $h: [0, 1] \rightarrow \mathbb{R}$ como

$$\begin{aligned} g_j(\lambda) &= \frac{L_x \nabla f_j(x)}{\|\nabla f_j(x)\|} \psi(\lambda) + \frac{1}{2} \|\psi(\lambda)\|^2, \quad \forall j \in \{1, 2\}, \\ \implies g'_j(\lambda) &= \frac{L_x \nabla f_j(x)}{\|\nabla f_j(x)\|} \psi'(\lambda) + \psi(\lambda)^T \psi'(\lambda), \\ h(\lambda) &= \max_{j \in \{1,2\}} g_j(\lambda). \end{aligned}$$

Então, o minimizador de h em $[0, 1]$ é $\tilde{\lambda}$. No caso em que $\frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} = \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|}$ então h é constante, logo, qualquer $\tilde{\lambda} \in [0, 1]$ gera a direção central, por exemplo, $\tilde{v} = \psi\left(\frac{1}{2}\right)$.

Suponha $\frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} \neq \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|}$.

Como g é convexa e diferenciável, $[0, 1]$ é convexo, e h é o máximo das funções componentes de g , então as condições do Teorema 3 estão satisfeitas. Esse teorema garante que se \tilde{v} minimiza h então $\exists \mu \in [0, 1]$ tal que duas condições são satisfeitas. A primeira é que

$$\begin{aligned} \mu g_1(\tilde{\lambda}) + (1 - \mu) g_2(\tilde{\lambda}) &= \max_{j \in \{1,2\}} g_j(\tilde{\lambda}) \\ \iff \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|} \psi(\tilde{\lambda}) + \mu \psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) &= \max_{j \in \{1,2\}} \frac{L_x \nabla f_j(x)}{\|\nabla f_j(x)\|} \psi(\tilde{\lambda}) \end{aligned}$$

$$\iff \mu \psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = \max\{0, \psi'(\tilde{\lambda})^T \psi(\tilde{\lambda})\}.$$

Portanto, se $\psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = 0$ então $\mu \in [0, 1]$.

A segunda condição do Teorema 3 é que, $\forall \lambda \in [0, 1]$,

$$\begin{aligned} & \left(\mu g'_1(\tilde{\lambda}) + (1 - \mu) g'_2(\tilde{\lambda}) \right) (\lambda - \tilde{\lambda}) \geq 0 \\ \iff & \left(g'_2(\tilde{\lambda}) + \mu(g'_1(\tilde{\lambda}) - g'_2(\tilde{\lambda})) \right) (\lambda - \tilde{\lambda}) \geq 0 \\ \iff & \left(\nabla f_2(x)^T \psi'(\tilde{\lambda}) + \psi(\tilde{\lambda})^T \psi'(\tilde{\lambda}) + \mu \|\psi'(\tilde{\lambda})\|^2 \right) (\lambda - \tilde{\lambda}) \geq 0 \end{aligned}$$

Como visto, se $\psi'(\tilde{\lambda})^T \psi(\tilde{\lambda}) = 0$ então qualquer $\mu \in [0, 1]$ pode ser tomado. Portanto, para isso ocorrer,

$$\begin{aligned} & \psi(\tilde{\lambda})^T \psi'(\tilde{\lambda}) = 0 \\ \iff & \left(-\frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} + \tilde{\lambda} \psi'(\tilde{\lambda}) \right)^T \psi'(\tilde{\lambda}) = 0 \\ \iff & \tilde{\lambda} = \frac{L_x \nabla f_1(x)^T \psi'(\tilde{\lambda})}{\|\nabla f_1(x)\| \|\psi'(\tilde{\lambda})\|^2}. \end{aligned} \quad (\text{D.6})$$

Note que

$$\begin{aligned} \|\psi'(\tilde{\lambda})\|^2 &= \left\| \frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} - \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|} \right\|^2 \\ &= \frac{2L_x^2 (\|\nabla f_1(x)\| \|\nabla f_2(x)\| - \nabla f_1(x)^T \nabla f_2(x))}{\|\nabla f_1(x)\| \|\nabla f_2(x)\|}, \end{aligned} \quad (\text{D.7})$$

e que

$$\begin{aligned} \nabla f_1(x)^T \psi'(\tilde{\lambda}) &= \nabla f_1(x)^T \left(\frac{L_x \nabla f_1(x)}{\|\nabla f_1(x)\|} - \frac{L_x \nabla f_2(x)}{\|\nabla f_2(x)\|} \right) \\ &= \frac{L_x \|\nabla f_1(x)\|^2}{\|\nabla f_1(x)\|} - \frac{L_x \nabla f_1(x)^T \nabla f_2(x)}{\|\nabla f_2(x)\|} \\ &= \frac{L_x \|\nabla f_1(x)\| (\|\nabla f_1(x)\| \|\nabla f_2(x)\| - \nabla f_1(x)^T \nabla f_2(x))}{\|\nabla f_1(x)\| \|\nabla f_2(x)\|}. \end{aligned} \quad (\text{D.8})$$

Juntando (D.7) e (D.8) em (D.6),

$$\tilde{\lambda} = \frac{L_x \|\nabla f_1(x)\|}{2L_x \|\nabla f_1(x)\|} = \frac{1}{2}.$$

Portanto, $\psi\left(\frac{1}{2}\right)$ é direção central. ■